Essays in antitrust economics

Verouden, V.C.H.M.

Publication date:
2001

Document Version
Publisher's PDF, also known as Version of record

Essays in Antitrust Economics

Essays in Antitrust Economics

Proefschrift

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Brabant, op gezag van de rector magnificus, prof. dr. F.A. van der Duyn Schouten, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op maandag 25 juni 2001 om 14.15 uur door

Vincentius Cornelius Henricus Maria Verouden

geboren op 19 augustus 1971 te Weert.

Promotor: prof. dr. E.E.C. van Damme

Acknowledgements

This dissertation has been shaped and influenced by the advice and support I have received from my supervisors, colleagues and friends. Here, I would like to acknowledge those people, whose help and support have been very valuable in helping me to complete this dissertation.

I started my PhD studies at the CentER for Economic Research of Tilburg University in September 1995. Tilburg is without doubt an excellent place for study and research. It has good facilities and, most importantly, fine people. Prof. Eric van Damme has supervised me throughout the whole period. He forced me to think more independently and confronted me with the many shortcomings in the research work that I submitted. Apart from supervising my thesis, he also involved me in a number of his consulting projects. Eric, thanks a lot for all your efforts.

During the whole of 1997 and again during several months in 1998-99, I had the opportunity to visit the research institute GREMAQ in Toulouse, as part of the ENTER exchange network (European Network for Training in Economic Research). The hospitality of GREMAQ is gratefully acknowledged. During this period, Prof. Patrick Rey was so kind as to supervise me, for which I am very grateful. I consider my stays at the Toulousian institute, with its top seminars, faculty and students, as the most inspiring periods of my academic life. The mere possibility to be together with other students who are interested in the same topics as you are is invaluable. As far as I can see, the fact that a number of PhD students have to share the same office is an asset, rather than anything else.

Of great impact on my PhD has been the work that I have done with Jérôme Pouyet from Toulouse. We started our research on cartel formation in the Autumn of 1997 and only recently finished a first paper (included as Chapter 6 of this thesis). Even though the work has taken quite some time and has been tough, it has proved to be very rewarding. This period has confirmed my belief that nothing can beat teamwork. Jérôme, thanks a lot. I hope we will get the opportunity to continue

academic work at some point in the future.

As of January 2000, I started to work as a consultant for the Merger Task Force of the European Commission. It has not been easy to combine the finishing of a PhD thesis with the requirements of such a job. Nonetheless, the good atmosphere in my work for the MTF has certainly helped me in keeping up the effort required for my thesis. Many thanks to my colleagues, fellow consultants and friends, in particular.

Besides my supervisors in Tilburg and Toulouse, I would also like to thank the members of the thesis committee: Stephen Martin, Frank Verboven, Sytse Douma and Leigh Hancher. Thank you for the time spent and the comments given. Many thanks also to Dirk Van Erps, F. Enrique Gonzalez-Diaz and Jérôme Pouyet for having commented on parts of this thesis. I am grateful to Jan Bouckaert, Paul de Bijl, Jos Jansen, Emiel Maasland, Sander Onderstal, Thibaud Vergé, Guido Friebel, Lucy White, Paul Seabright, Patrick Bolton and Christof Swaak for a number of interesting occasions for discussion. Equally, I would like to thank the secretaries and staff of CentER and GREMAQ for their administrative support.

On a personal level, I can only be thankful to many people who have supported me all along. Special thanks go out to my father and mother; they are wonderful people (on top of this, my father has read the entire thesis and corrected the English text). Gabriëlle, Willian, Simone, Astrid, Marijn, Stan and Patries: I am happy that you are close to me as well. Thank-yous also to the vaguely defined *Vijf Vrolijke Vrienden*: Nils, Werner, Wessel, Kees, Alwin and all the others in the entourage. Finally, Jérôme, Sonia, Khaled, ... thanks for the good friendship. I hope that I will now again have lots of time to spend with all of you !

Brussels, March 2001.

Contents

Chapter 1

Introduction

Competition law or *antitrust law,* as it is called in the United States, is a field of law for which economic concepts are of central importance. Many of the key concepts of competition law - for example, 'competition', 'restriction of competition', 'anti-competitive effect' - are concepts which are clearly economic by nature, if not even rooted in economics. This is not to say, however, that the reasons for adopting competition laws have always been strictly 'economic', or that always purely economic interpretations have been given to these concepts. Rather, political, social and even moral considerations have been at the forefront in many instances[1].

The adoption process of the earliest antitrust laws in the United States provides a good example[2]. In the late 19th century, the U.S. economic landscape exhibited a strong consolidation process in the form of *trusts*, legal arrangements by which owners of different companies transferred their control to a trustee in return for trust certificates entitling them to a proportionate share of the profits in the jointly managed companies. The Standard Oil Trust, for example, controlled about 95% of the oil production in the US. Similar concentrations were not uncommon in banking, the railways, tobacco and other sectors. This situation did not leave the Americans indifferent. The end of the 19th century was marked by deep social unrest and economic uncertainty. In the public opinion, the well-publicized vested interests and special priviliges of the trusts and other big business concentrations were the root cause of these problems. Commentators of the time even labelled the trusts 'conspiracies',

[1] This apposition is perhaps more apparent than real. After all, one can argue that the political, social and moral considerations were at least in part reflective of the prevailing distribution of wealth and influence in society.

[2] Cf. Hovenkamp (1999), Jenny (1993), Fasquelle (1993), Rubin (1976).

exercising a 'demoralising influence'[3]. Faced with strong popular dissatisfaction with the trusts, the U.S. Congress adopted its first antitrust law, the Sherman Act, in 1890.

Economists, however, hardly played a role in the realisation of the Sherman Act. In part, this may be explained by the academic position of late 19th century American economists: they generally favoured a rather dogmatic *laissez-faire* approach and therefore did not feel comfortable with government intervention in the first place. In addition, they perhaps lacked the tools to give a clear description of the central antitrust issues as well: it is illustrative that the members of the American Economic Association, created in 1885, did not consider it necessary to intervene in the political debate to influence the Congress or the contents of the Sherman Act[4]. In this sense, the adoption of the Sherman Act was a choice advocated by politicians, rather than a choice inspired by economists.

It was to take a while for economics to take a more important role in the application of the Sherman Act. Section 1 of the Sherman Act states that 'every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States (...) is hereby declared to be illegal'. In the landmark case *Board of Trade of Chicago* of 1918, the U.S. Supreme Court set the test which has since become standard[5]. In my view, it is difficult to come up with a better description of the role of economics in antitrust policy than the one given by this Court:

> 'The true test of legality is whether the restraint imposed is such as merely regulates and perhaps thereby promotes competition or whether it is such as may suppress or even destroy competition. To determine that question, the court must ordinarily consider the facts peculiar to the business to which the restraint is applied; its condition before and after the restraint was imposed; the nature of the restraint and its effect, actual or probable. The history of the restraint, the evil believed to exist, the reason for adopting the particular remedy, the purpose or end sought to be attained, are all relevant facts. This is not because a good intention will save an otherwise objectionable

[3]Morgan , W. (1889) 'History of the Wheel and Alliances and Impending Revolution', p.15. Quoted in Fasquelle (1993).

[4]Cf. Greenhut and Benson (1989).

[5]*Board of Trade of Chicago v. United States* 246 U.S. 231 (1918).

> regulation or the reverse; but because knowledge of intent may help the court to interpret facts and to predict consequences'.

It is by no means a simple matter to determine from the outset whether a given contract or business practice is good or bad for competition. For example, if two competing companies enter into a co-operation agreement for the development of a new technology, what will be the effect on competition in the market? If the two companies are of a modest size, probably not much harm is to be expected, rather the contrary. On the other hand, if the two companies are each other's closest competitors, their co-operation may well lead to a substantial reduction of competition in the market. But where to draw the line? Similarly, if the largest manufacturer in the industry concludes an exclusive distribution contract with the strongest distributor, what must we think of this?

In the first place, as suggested by the Supreme Court's ruling, it is only by considering the specifics of the case that sound conclusions can be drawn concerning the competitive nature of a given contract or a business practice. Further, this will often require not just an examination of the possible effects, but also an inquiry into the possible motivations of companies to adopt such market behaviour. After all, in view of the inherent difficulties in evaluating and predicting the effects of most kinds of market behaviour, it must be instructive to complement the analysis by exploring *why* rational companies in a competitive situation choose to act in such a way: 'knowledge of intent may help the court to interpret facts and to predict consequences'[6]. If the assessment of the possible effects is inconclusive, an analysis of the business motivations may shed sufficient light on the case.

In economics, it is the field of Industrial Organization that studies the behaviour of firms and its implications for the functioning and structure of markets. This field was initially empirical by nature, focusing on obtaining descriptive statistics and correlations among industry variables. Very well known is the resulting 'structure-conduct-performance' paradigm, developed by economists from the University of Harvard (Edward Mason, Joe Bain and others). According to this paradigm, market structure (the number of firms in the market, the degree of vertical integration and so on) determines market behaviour (prices, investment in R&D, advertising, ...), which in turn results in performance (efficiency, profits). However, with its emphasis on finding empirical

[6] *Board of Trade of Chicago v. United States* 246 U.S. 231 (1918).

regularities, it did not give much insight into the underlying 'mechanics' of markets nor into the rationality of some of the observed market behaviour.

Since the 1970s, Industrial Organization has complemented (and in some cases turned upside down) the traditional empirical line of research with a more rigorous theoretical analysis, focusing, among other things, on the rationality of market behaviour. An important tool in this line of research has proved to be game theory. This theory studies strategic or competitive interaction using mathematical models[7]. A game model specifies the players in a game (for example, firms in a market or individuals in an organisation), the information they have (or do not have), the actions they can choose, the timing of these actions, the pay-offs for each player that result from the actions which are chosen and the preferences of the players over the possible pay-offs. In such a model, each player is supposed to choose a strategy (a plan of action) that maximizes his pay-offs (or, more generally, his utility level) based on the information available to him and his expectations about his rivals' actions. The widely accepted solution concept to game models is the so-called Nash equilibrium (Nash, 1951). This solution concept represents an equilibrium in the sense that each player maximizes his pay-offs while correctly anticipating what the other players are going to do[8].

For the greater part, this thesis will be about the application of such game-theoretic analyses to vertical agreements, i.e. agreements concluded between firms operating at different levels of the production or distribution chain, such as between suppliers and retailers. Many relationships between suppliers of goods and their distributors go well beyond simple agreements to deliver goods at a certain unit price. Often, these relationships are governed by medium or long term contracts that impose certain obligations on one or both parties, restricting in some way their commercial freedom. For example, a supplier who grants an exclusive sales territory to a distributor necessarily commits itself not to sell to other distributors based in that specific area. Contractual obligations of this kind are commonly referred to as 'vertical restraints'.

[7] A good and comprehensible survey of game theory - as it relates to the law - can be found in Baird, Gertner and Picker (1994). A more formal introduction to game theory is provided by Gibbons (1992).

[8] For a discussion on the precise relation between rationality and the use of an equilibrium concept, see Mas-Colell e.a. (1995).

Overview of the thesis

The thesis is divided into two parts. The first part, consisting of Chapters 2 and 3, has a review character. Chapter 2 reviews the economic literature on the competitive effects of vertical agreements. The purpose of the chapter is to obtain an understanding of the main insights that economic theory has provided as regards such agreements. Chapter 3 analyses the role of economic analysis from a legal perspective, by focusing on the role of economic analysis in the application of the competition rules of the European Union towards vertical agreements. As will transpire from these two chapters, European competition policy has its peculiarities but is more and more moving towards a real balancing of pro- and anti-competitive effects.

The second part of the thesis will provide three concrete applications of game-theoretic analyses. In two chapters, Chapter 4 and Chapter 5, the rationality of the use of a particular type of vertical restraint, namely resale price maintenance, will be tested in specific market circumstances. Under resale price maintenance, a manufacturer requires retailers not to sell its products below a certain minimum price[9]. It will follow from these chapters that the need for retailers to recover their fixed costs plays a determining role in the type of vertical restraint that will be used by the manufacturer. Finally, Chapter 6 is about cartel formation in industries where firms are uncertain about each other's cost levels (and, therefore, about each other's pricing incentives). In a specific model, the general conception will be tested that the likelihood of firms forming a cartel is greater in concentrated industries than in industries with many firms.

A more detailed description of each of the Chapters 2 to 6 is set out below.

Chapter 2: A focus on vertical agreements

In the economic science, there have been quite a few shifts in attitude as to the admissibility of vertical agreements. The most notable shift occurred in the 1960s, a shift commonly associated with scholars from the University of Chicago, such as Aaron Director, Lester Telser, George Stigler and Robert Bork[10]. The 'Chicago School'

[9]The analysis also applies to related practices. For example, a manufacturer's unilateral policy not to deal with 'discounters' can be considered a kind of resale price maintenance.

[10]The Chicago School developed already as of the 1930s but it was only since the 1960s that it began to get truly influential, in economics and beyond (Reder, 1982). It should be noted that

stimulated a line of research with a rigorous focus on the motivations of companies using the restraints: why do rational companies in a competitive situation choose to be restricted in their choice possibilities? The aim was to find explanations for observed business conduct in line with the starting points of neoclassical theory (utility maximisation by rational economic agents) and consistent with the idea that market participants are generally capable of correcting and internalising possible market imperfections themselves[11]. In addition, the Chicago economists proposed to use the criterion of economic efficiency (welfare) as the sole normative standard against which the lawfulness of a given business practice should be tested[12].

The Chicago School emphasized that agreements concluded by companies in a vertical relationship are, by their nature, very different from agreements concluded by firms which are in direct competition with each other (also called 'horizontal agreements'). The fact that the former are agreements concluded by companies which each perform an indispensable function in putting the product on the market, suggests that they are primarily used to make the vertical combination more efficient. After all, in a vertical relationship one party will be damaged when the other party does not function properly. And 'properly' means, in by far most cases: from the point of view of the consumers, because in the end, they are supposed to buy the product. Through this special interdependent relationship, every party in a vertical agreement can, in principle, be considered a natural ally of the consumer. The permissive attitude towards vertical restraints has become widely known as the 'Chicago view'.

During the 1980s and 1990s, the Chicago methodology of studying the rationale of observed behaviour on the basis of rigorous theoretical analysis (and its emphasis on the use of an economic efficiency criterion to evaluate the impact of such behaviour) has gone to the centre ground of Industrial Organization[13]. Its sharpest conclusions (the 'Chicago view'), however, have not. For a large part this can be attributed to the increasing use of game theory in Industrial Organization, which allowed for

---

the focus of the University of Chicago was much wider than just Industrial Organization. In fact, 'Chicago economics' refers to the application of strict neoclassical theory to a great number of fields of study, such as the macro economy (Milton Friedman, Robert Lucas), the political process (George Stigler), sociological phenomena (Gary Becker), and the legal system (Richard Posner).

[11] See Reder (1982) for a detailed account of the assumptions made by the 'Chicago School'.

[12] Welfare can be conceived as the (weighted) sum of *consumer surplus* (a monetary measure of the utility derived from consumption) and *producer surplus* (such as profits). Normally speaking, the consumer surplus goes up when prices go down, when consumption levels go up, when the quality of products gets better, etc.

[13] Cf. Van den Bergh (1997).

the construction of a wider range of models and, correspondingly, a wider range of outcomes. These models suggested that vertical integration or contractual restraints could be rational and effective ways to engage in anti-competitive behaviour. For example, it was shown that delegating pricing decisions to exclusive distributors might allow producers to credibly commit to less competitive behaviour towards each other, making use of the fact that the incentives to compete on the distribution level differ from those on the producer level[14]. Similarly, some exclusive dealing contracts were shown to be possible tools for foreclosing markets, in particular because they render the anti-competitive objective (foreclosure) more credible and time-consistent[15].

As to the characterisation of the circumstances in which vertical restraints are likely to have positive or, on the contrary, detrimental effects for competition and welfare, the current body of economic literature offers fairly extensive material on which to base such a characterisation. The purpose of Chapter 2 is to provide an overview of the main insights that economic theory has provided. Rather than giving a long enumeration of the different market situations that have been studied and the corresponding results, the aim is to present and develop the main arguments, occasionally with the use of some simple examples and models.

Chapter 3: Vertical agreements and Article 81(1) EC

In the European Union a lively debate has taken place in recent years concerning the approach that should be adopted in competition policy towards vertical agreements. Among other things, this debate has centered around the question what role economic analysis should play in the application of Article 81 of the EC Treaty and, in particular, in the application of the first paragraph of the article, Article 81(1), which establishes the principle that agreements which are restrictive of competition (and which affect trade between member states) are prohibited.

Central to the application of Article 81 EC is the notion of what constitutes a 'restriction of competition' under Article 81(1). While Article 81(1) states a few broad examples of what might constitute such restrictions, these have left ample opportunity for different interpretations in the application of this article in individual cases, notably in the context of vertical agreements. The European Commission, the central institution in the application of Article 81, has tended to be rather strict in its

[14] Cf. Rey and Stiglitz (1995).
[15] Cf. Aghion and Bolton (1987).

interpretation, in the sense that many vertical agreements were seen as constituting restrictions of competition, given that they reduce the commercial freedom of the contracting parties and of third parties. Throughout the years, there have been regular calls for a more economics based analysis of the concept 'restriction of competition' under Article 81(1)[16]. According to the critics, the Commission's approach was too formalistic and, as a result, inefficient in making a distinction between the competition enhancing effects of vertical agreements and the effects restricting competition - pre-eminently a matter of economic analysis.

The purpose of Chapter 3 is to obtain an understanding of the main developments in the interpretation of the concept 'restriction of competition' in Article 81(1) in relation to vertical restraints. The focus will be on the developments in the jurisprudence of the European Courts[17], as it is the most important source of guidance.

It emerges from more than 30 years of jurisprudence that the words 'restriction of competition' have been interpreted in the light of the overall objectives of the EC Treaty (in particular, the creation of a single European market), rather than in the light of competition as such. Nonetheless, it is striking to see how many points of reference particularly the jurisprudence of the European Courts has offered for a greater role for economic analysis, notably under Article 81(1). On a lighter tone: the analysis also shows that attempts by economists to comment on concepts which appear 'economic' at first sight, need not always lead to comments which are precisely to the point. This in particular applies to commenting on the concept 'restriction of competition' in Article 81(1) without taking into account the role of Article 81(3), the exemption possibility to the rule of prohibition, or the particular attribution of competences in the enforcement system of the competition rules.

Whereas the role of economic analysis in the application of Article 81 has by no means been a constant one throughout the years, in recent years it can be said to evolve very rapidly. On a policy level, the main driver behind this development has been the publication of the Green Paper (a consultation document) of the European

[16]Economists typically mean by this an analysis to establish whether or not particular agreements reduce (consumer) welfare: only when an agreement reduces welfare, should it be deemed 'restrictive of competition'. Some of the critics may have been referring to more 'limited' forms of economic analysis (e.g. regarding the question whether or not a restriction is necessary for the attainment of some commercial objective; see Chapter 3).

[17]The European Court of Justice and, since 1988, also the Court of First Instance. On appeal, these institutions review the decisions adopted by the European Commission in the application of Article 81. In addition, the Court of Justice gives rulings on questions of law asked by national courts concerning the interpretation of Article 81.

Commission concerning vertical agreements in 1997, in which the subject of economic analysis took a prominent place[18]. Since then, a number of steps have been taken to reform EU competition policy towards vertical restraints. On the substantive side, these steps can be seen as increasing the role for economic analysis in the application of Article 81 as a whole. In this respect, the integral application of the whole of Article 81 by the Commission, national Courts and national competition authorities (as proposed by the most recent policy initiative) appears to me to be a logical and welcome step. There remain a number of questions, however, such as the question who, in practice, will bear the burden of proof in cases under Article 81. While the principle is clear (the burden of proving an infringement of Article 81(1) rests on the party alleging the infringement, a party claiming the benefit of Article 81(3) shall bear the burden of proving that the conditions of that paragraph are fulfilled), much will depend on the Commission's own style in characterising the affected markets and in assessing vertical restraints, whether the burden will be at the requisite level.

Chapter 4: Resale price maintenance in a spatial market

Chapter 4 is the first chapter in the thesis providing for a concrete game-theoretic analysis. It applies to the use of resale price maintenance. Under this practice, a manufacturer obliges retailers not to sell its products below a certain minimum price. Likewise, a manufacturer's unilateral policy not to deal with 'discounters' can be considered a kind of resale price maintenance.

As will be discussed in Chapter 2, several explanations for the use of minimum resale price maintenance (RPM) have been given in the economic literature. One explanation that has received little attention recently is the *outlets hypothesis*, articulated by Yamey (1954) and elaborated upon by Gould and Preston (1965). The outlets hypothesis assumes that final demand for the manufacturer's product is a function both of the retail price and the number of retail outlets: the price-demand schedule for a product shifts outward if the number of retailers carrying the product increases. One of the informal arguments for this positive relationship given is that the inconvenience of shopping (e.g. travelling) is reduced when retail density is higher. Gould and Preston then argue that price floors, by raising the retail margin above the competitive level, lead to an increase in the number of retail outlets in a

[18]Green Paper on Vertical Restraints in EC Competition Policy, 22 January 1997, COM(96) 721 final.

free-entry market equilibrium and, because of the positive impact of this increase on final demand, to higher profits for the manufacturer.

In determining the optimal number of retail outlets, the manufacturer has two opposite effects to take into account. First, the higher the number of retail outlets entering the market in equilibrium, the higher the sum of fixed costs involved (to be covered through the retail mark-up). On the other hand, an increase in the equilibrium number of retailers may go with a decrease in the 'effective price' faced by consumers as the average travelling distance for consumers decreases. For a given wholesale price, a drop in the 'effective price' benefits the producer as the total quantity of goods sold increases. Therefore, the reduction in transportation costs incurred by the consumers may allow the producer to capture a larger part of the consumer surplus.

Two articles which have evaluated the two above effects in a context of spatial retail competition[19], Mathewson and Winter (1983) and Bittlingmayer (1983), find that the first effect - as evaluated at the market equilibrium without vertical restraints - appears to always dominate the second effect. The main reason for this result is that in the absence of vertical restraints, there is a strong 'business stealing effect' (Tirole, 1988): retailers, when deciding to enter the market, do not take into account the negative effect of entry on the profits of the other retailers. From the viewpoint of the industry, this leads to a certain bias towards excess entry. It also renders the result that price *ceilings*, and not price *floors*, are required if the producer wants to maximize profits by influencing the number of retail outlets in the long run.

In Chapter 4, I will verify whether the above described results for the outlets hypothesis are due to the particular transportation cost assumptions of the underlying models. In line with the majority of spatial models of retail competition, these models have made the assumption that transportation costs depend proportionally on the quantity of products actually bought at the retail outlet. By contrast, I will assume that customers only incur a fixed cost when visiting a retail outlet, i.e. a cost irrespective of whether they buy several products or nothing at all. This assumption seems quite justified when the size or quantity of the goods transported is relatively small or transportation costs are looked upon as opportunity costs associated with the time spent on shopping and not on other activities.

An interesting feature of the fact that transportation costs are not dependent on

[19]Spatial retail competition refers to competition between retailers that are differentiated through distance. Retailers being differentiated amounts to retailers possessing some (localized) market power.

the quantity bought is that it changes the analysis of where a consumer will buy. The larger the quantity consumers want to buy, the more they will get interested in going to a remote shop with a low price, rather than going to the shop closeby with high prices. In a way, the choice of where to buy also depends on the choice how much to buy. The fixed transportation cost structure thus appears to bring about more competitive retail conditions than the proportionally dependent transportation cost structures. In principle, this may reduce the extent of the excess entry bias of spatial models and, therefore, lead to price floors being more attractive as a means to foster entry by retailers.

However, despite this feature it is found that also with the fixed transport cost specification, the manufacturer does not find it profitable to impose price floors. Again, the better capture of consumer surplus appears not to weigh up against the increase in fixed costs involved with the larger retail network. To put the conclusion of this analysis into business motivations in the negative: if one observes a company employing price floors in a setting such as the one studied in Chapter 4, this company is not doing it to increase its retail network. Rather, other motivations (anti-competitive motivations?) must be underlying its use.

Chapter 5: Resale price maintenance under cost uncertainty

As will be discussed in Chapters 2 and 4, several explanations for the use of resale price maintenance have been given in the economic literature. The emphasis of Chapter 5 is on the incentive and insurance properties of resale price maintenance in an uncertain trading environment, a subject that has notably been analysed by Rey and Tirole in their article 'The logic of vertical restraints' (1986, American Economic Review 76: p. 921-936). In this article, Rey and Tirole set up a spatial model of retail competition and analyse the role of resale price maintenance (RPM) when there is uncertainty about future demand and cost levels. The retailers are assumed to be better informed (ex post) about the realisation of final demand and about their own costs than the manufacturer.

The basic trade-off in the choice of contract is between the optimal exploitation of market power and the amount of risk that the retailers are willing to accept. The optimal exploitation of market power requires that one avoids the double marginalisation problem associated with linear wholesale pricing, i.e. the problem that final prices end up too high from the viewpoint of the vertical chain due to consequent stages of

market power (Spengler, 1950). One way of doing so is to use two-part tariffs, involving a fixed upfront payment (e.g. a franchise fee) and a low marginal wholesale price. Another method is resale price maintenance, by which the producer directly imposes the proper final price on the retailers. If there is no uncertainty, the two methods yield identical results. However, when there is uncertainty the two may differ. Whereas free competition between retailers clearly allows better use of local information than the inflexible instrument of RPM, the insurance properties are more complex: risk averse retailers dislike the burden of a franchise fee, but also RPM exposes them to risks, to the extent that their profit margins are subject to fluctuations.

One of Rey and Tirole's specific results is that when there is uncertainty about *costs*, free competition between retailers (in combination with a two-part tariff) has good incentive and insurance properties in the various cases considered in the article[20]. The principal reason is that under cost uncertainty, the retail margin is particularly volatile under RPM: the retail price is fixed but the cost level varies, a feature for which risk averse retailers need some compensation (e.g. via lower wholesale prices).

The principal aim of Chapter 5 is to show that Rey and Tirole's result concerning the favourable incentive and insurance properties of retailer competition does not generally carry over to the case in which retailers possess some degree of market power (due to retailer differentiation) and face retailer-specific cost uncertainty. When differentiated retailers face firm-specific risks on cost levels, resale price maintenance may be a more profitable instrument for the manufacturer than free retail competition.

The essential point is that in the absence of market power associated with retailer differentiation, there is only one source of double marginalisation: the double marginalisation due to cost differences at the retail level (when one retailer turns out to be more efficient than the other, it obtains a positive retail margin by just undercutting the price of the other retailer). With only this source of double marginalisation, the need for a powerful two-part wholesale tariff is not very great: even when retailers are so risk averse that they are not willing to accept any positive franchise fee at all (because this might result in a loss in some situations), free downstream competition

[20]Free retailer competition turns out to be a more profitable option for the manufacturer than RPM in all three model specifications analysed: 1) the case in which cost uncertainty is market-wide (the costs faced by all retailers fluctuate in the same way) and retailers have no market power 2) the case in which cost uncertainty is market-wide and retailers derive market power from being differentiated 3) the case in which cost uncertainty is firm-specific ('idiosyncratic') and retailers lack market power unless they are more efficient than the competition.

turns out to perform better than resale price maintenance[21].

However, when there is also a double marginalisation problem due to retailer differentiation, there is a greater need for a powerful wholesale tariff, involving lower wholesale prices and, correspondingly, a higher franchise fee. But the extent to which the manufacturer can charge the required franchise fee is limited by the risk that the retailers are willing to bear: a franchise fee can only be recovered by the retailers when the retail margin they earn is sufficiently positive (in some average sense)[22]. Whereas firm-specific cost uncertainty makes it more difficult to charge a franchise fee to risk averse retailers which are competing in prices, resale price maintenance is an instrument to protect the retailers against more efficient rivals. As a result of this insurance property, resale price maintenance can be an optimal commercial policy for a manufacturer. In particular, it is shown that in the case of differentiated retailers, RPM is optimal when the cost uncertainty is firm-specific, the retailers are sufficiently risk averse and the range of possible retail cost levels is not too wide (so as to make RPM too 'rigid' as an instrument).

To put the conclusion of this analysis into business motivations in the positive: if one observes a company employing price floors in a setting such as the one studied in Chapter 5, this company may well be doing it because it is in its own interest to better protect its retailers.

Chapter 6: Cartel formation under incomplete information

It is generally thought that the likelihood of firms forming a cartel is greater in concentrated industries than in industries with many firms. Not only because it is, so the argument goes, easier to monitor a cartel agreement in the relatively surveyable environment of a tight oligopoly (cartel enforcement argument) but also because it may be easier or more attractive for fewer firms to come to terms about the conditions applying to the cartel (cartel formation argument)[23].

One element that can be a source of difficulty in the formation of cartels is the problem of *incomplete information* with respect to the cost levels of the participating

[21] Cf. Rey and Tirole (1986). When no franchise fee is imposed, the insurance properties of direct competition are necessarily very good: even when the retail margin ends up being zero (which occurs whenever the competitors are at least as cost efficient), the retailers just break even.

[22] In a similar fashion, the results of Rey and Tirole (1986) also depend on the specific assumption made as regards the level of fixed costs required to set up a retail outlet.

[23] See Phlips (1995) for an overview of the literature.

firms. This information asymmetry may pose a problem at the stage where the cartel must determine the conditions of the cartel agreement (e.g. production quotas, fixed market shares) for the participants. Obviously, the conditions of the cartel agreement also bear on the decision to join the cartel in the first place. For example, a firm which is relatively efficient will typically only be satisfied with a production quota that somehow steps up to this fact (in other words, the quota must be relatively large), otherwise it may just prefer to compete on the market. This, however, should induce firms which are less efficient to overstate their efficiency in order to obtain a higher share of the cartel output. But when every firm is saying to be efficient (or saying to have become more efficient since the latest negotiations) and claiming large quota, this reduces the attractiveness of the cartel for firms which are effectively among the most efficient.

The extent to which cartel agreements can overcome the conflicting requirements mentioned above has been the subject of extensive research[24]. In order to characterise the outcomes that a cartel can achieve in situations of incomplete information, the issue of cartel formation has typically been approached using a standard mechanism design approach: in an industry, there is a 'cartel manager' who proposes a cartel arrangement and determines the optimal quotas depending on the costs each firm announces to have[25]. Given this scheme, firms decide whether or not to join and, if they do, they announce their costs. A proposed cartel agreement is called 'efficient' when only the firm(s) with the lowest cost produce(s). In order to form such a cartel, the cartel manager must, according to the well-known Revelation Principle, propose a scheme (possibly involving side payments) that ensures individual participation and induces the firms to reveal their cost information.

The current literature does provide some justifications for the generally held belief that cartels are most difficult to form in industries with many firms, but it fails to do so in several constellations. In particular, in a setting where the number of possible efficiency types is limited to two (firms are either efficient or inefficient), Kihlstrom and Vives (1989, 1992) have shown that the formation of an efficient cartel *is* possible,

[24] Cf. Cramton and Palfrey (1990) and Kihlstrom and Vives (1989, 1992).

[25] In studying the problems related to cartel formation, the literature abstracts away from the cartel enforcement problem (and vice versa). Cartel arrangements are assumed to be enforceable, even though the secrecy of the arrangement implies that there is, for example, no public authority available to enforce it. The assumption of enforceability is a short-cut to capture in a static context the reputation of the cartel manager and the firms participating in the cartel which guarantee the self-enforceability of the arrangement in a dynamic context (e.g. by means of trigger strategies).

both in the case of a duopoly and in the borderline case of an industry comprising very many firms (modelled as a continuum of firms). The reason of this latter, rather surprising, result is that in such an industry there is little uncertainty about the type of firm that should produce nor about the fraction of efficient firms being present, so that it turns out to be not so difficult to reconcile all individual participation and incentive requirements. This result holds for all meaningful probability distributions on the two cost types.

The purpose of Chapter 6 is to consider again the issue of cartel formation in the above context, but from a different angle: in order to characterise the possible outcomes that a cartel can achieve, it is proposed to explore the additional requirement of *collusion-proofness*. The above mentioned models of the cartel manager (the principal) trying to obtain the efficient cartel outcome by inducing truthful cost revelation by the firms (the agents), all use the standard assumption that every agent behaves non-cooperatively: no communication is possible between the agents, which is a standard assumption for the Revelation Principle. The aim of Chapter 6 is to see whether the obtained results continue to hold when communication between groups of firms *cannot* be excluded and, in particular, when groups of firms try to (secretly) *coordinate* their cost announcements in order to obtain a better result.

The possibility of collusion by subcoalitions is shown to change the set of implementable rules, but *not* to change the principal result that efficient cartel formation is possible for any number of firms in the industry. Similarly, cases can be identified in which the extra requirement of collusion-proofness need not have an impact on the minimal level of transfers that is required to form these cartels. It appears that these results are due to a range of factors. An important one is undoubtedly the strong congruence of interest between the cartel manager and the individual cartel members: after all, the cartel manager is acting costlessly on behalf of the members, by maximizing their total expected profits. The fact that the cartel manager can freely use side transfers is a relevant aspect as well. With restrictions on the use of side-transfers, efficient cartels may not always be possible in the first place.

It emerges from this chapter that while the literature on cartel formation provides some theoretical support for the idea that cartels are most difficult to form in industries with many firms, it still leaves a number of open ends.

References

Aghion, P. and Bolton, P. (1987), 'Contracts as a barrier to entry', American Economic Review, Vol. 77: 388-401.

Baird, D., Gertner, R. and Picker, R. (1994) 'Game theory and the law' Harvard University Press, Cambridge, Massachusetts, USA and London, UK.

van den Bergh, R. (1997) 'Economische analyse van het mededingingsrecht' Rechtseconomische Verkenningen, Gouda Quint - Kluwer, Deventer, Netherlands.

Bittlingmayer, G. (1983) 'A model of vertical restriction and equilibrium in retailing' Journal of Business 56(4): 477-498.

Cramton, P. and Palfrey, P. (1990) 'Cartel enforcement with uncertainty about costs' International Economic Review, 31: 17-47.

Fasquelle, D. (1993) 'Droit américain et droit communautaire des ententes: étude de la règle de raison' Joly Editions, Paris, France.

Gibbons, R. (1992) 'A primer in game theory' Harvester Wheatsheaf, New York/London.

Greenhut, M. and Benson, B. (1989) 'American antitrust laws in theory and in practise' Avebury.

Gould, J. and Preston, L. (1965) 'Resale price maintenance and retail outlets' Economica 32: 303-12.

Hovenkamp, H. (1999) 'Federal antitrust policy: the law of competition and its practice', Hornbook Series, West Publishing Co., St. Paul, Minnesota, USA.

Jenny, F. (1993) 'Competition and efficiency' in Hawk (ed.): Annual Proceedings of the Fordham Corporate Law Institute 1993 - International Antitrust Law & Policy, Fordham Corporate Law Institute, New York, USA.

Kihlstrom, R. and Vives, X. (1989), 'Collusion by asymmetrically informed duopolists', European Journal of Political Economy, 5: 371-402.

Kihlstrom, R. and Vives, X. (1992), 'Collusion by asymmetrically informed firms', Journal of Economics and Management Strategy, 1: 371-396.

Mas-Colell, A., Whinston, D., Green, J. (1995) 'Microeconomic theory' Oxford University Press, Oxford/New York.

Mathewson, G. and Winter, R. (1983) 'Vertical integration by contractual restraints in spatial markets' Journal of Business 56(4): 497-526.

Nash, J. (1951) 'Non-cooperative games' The Annals of Mathematics 54: 286-295.

Phlips, L. (1995) 'Competition policy: a game-theoretic perspective', Cambridge University Press, Cambridge, United Kingdom.

Reder, M. (1982) 'Chicago economics: permanence and change' Journal of Economic Literature 20: 1-38.

Rey, P. and Tirole, J. (1986) 'The logic of vertical restraints' American Economic Review, 76: 921-39.

Rey, P. and Stiglitz, J. (1995) 'The role of exclusive territories in producers' competition' Rand Journal of Economics, 26(3): 431-451.

Rubin, S. (1976) in 'Competition law in Western Europe and USA' USA, Volume 3a, Gijlstra.

Spengler, J. (1950) 'Vertical integration and antitrust policy' Journal of Political Economy 58: 347-352.

Selten, R. (1973) 'A simple model of imperfect competition where four are few and six are many', International Journal of Game Theory, 2: 141-201.

Tirole, J. (1988) 'The theory of Industrial Organization', MIT Press, Cambridge, Massachusetts, USA.

Yamey, B. (1954) 'The economics of resale price maintenance' Sir Isaac Pitman, London, UK.

Chapter 2

Vertical agreements

2.1 Introduction

Many relationships between suppliers of goods and their distributors (e.g. wholesalers and retailers) go well beyond simple agreements to deliver goods at a certain unit price. Often, these relationships are governed by medium or long term contracts that impose certain obligations on one or both parties, restricting in some way their commercial freedom. For example, a supplier who grants an exclusive sales territory to a distributor necessarily commits not to sell to other distributors based in that specific area. Similarly, distributors who want to become part of a franchise network usually have to commit not to disclose the know-how that they receive in the context of the franchise agreement to companies outside the network. Contractual obligations of this kind are commonly referred to as 'vertical restraints'[1]. Furthermore, contracts between suppliers and distributors frequently involve rather elaborate payment schemes, such as quantity discounts, fixed fees or royalties[2].

During the 20th century, there have been quite a few shifts in attitude as to the admissibility of such restraints. Leading decisions early in the development of antitrust law reflected the somewhat orthodox view that restraints of all sorts reduce the economic freedom to act of the trading parties and that they are, hence, bound to

[1] At the end of this chapter, a short glossary can be found with descriptions of the main types of vertical restraints.

[2] In the economic literature, the mentioned payment schemes are sometimes also referred to as 'vertical restraints'. As the more general payment schemes are not, by themselves, restricting the actions that one of the parties can take, it may be preferable to distinguish between the two (cf. Katz, 1989).

interfere with the free play of trade. In particular, they would lead to interrupted seller access to customers and should thus be considered bad for competition[3]. In this respect, there was hardly any difference between the treatment of horizontal agreements (agreements concluded between firms operating at the same level of the production or distribution chain, i.e. competing firms) and vertical agreements (between firms operating at different levels of the production or distribution chain). After World War II, this approach found support in a number of empirical studies that tended to show a positive relationship between dense market structures and price and profit levels in the industry[4].

Since the 1960s, a line of thought commonly associated with the University of Chicago, has changed the direction of the debate[5]. The 'Chicago School' stimulated a line of research with a rigorous focus on the motivations of companies using the restraints: why do rational companies in a competitive situation choose to be restricted in their choice possibilities? The aim was to find explanations for observed economic behaviour in line with the starting points of neoclassical theory (utility maximisation by rational economic agents) and consistent with the idea that market participants are generally capable of correcting and internalising possible market imperfections themselves[6]. In addition, the Chicago economists proposed to use the criterion of economic efficiency (welfare) as the sole normative standard against which the lawfulness of a given business practice should be tested[7]. According to Bork (1978), 'Antitrust policy cannot be made rational until we are able to give a firm answer to one question: What is the point of law - what are its goals? Everything else follows from the answer we give'.

The Chicago School emphasized that agreements concluded by companies in a

[3]For the US, see Brennan (2000) or, in the context of the American 'rule of reason', Chapter 3. For the European Community, see Chapter 3.

[4]Fasquelle (1993, p. 46) notes that this positive relationship was perceived to hold for horizontally concentrated as well as vertically integrated industries.

[5]Cf. Brennan (2000). The Chicago School developed already as of the 1930s but it was only since the 1960s that it began to get truly influential, in economics and beyond (Reder, 1982).

[6]See Reder (1982) or Posner (1979) for a detailed account of the assumptions made by the 'Chicago School'.

[7]Total welfare can be conceived as the (weighted) sum of consumer surplus (a monetary measure of the utility derived from consumption) and producer surplus (such as profits). The weights put to consumer surplus and producer surplus imply a certain value judgment. The Chicago School proposed that the weights should be taken equal, rendering antitrust policy 'nonpolitical', i.e. not taking sides in any political dispute about how wealth ought to distributed among interest groups (Hovenkamp, 1999).

vertical relationship are, by their nature, very different from agreements concluded by firms which are in direct competition with each other (also called 'horizontal agreements'). The fact that the latter are agreements concluded between companies in a vertical relationship - that is, between companies which both fulfil an indispensable function in putting the product on the market - suggests that they are primarily used to make the vertical combination more efficient. After all, in a vertical relationship one party will be damaged when the other party does not function properly. And 'properly' means, in by far most cases: from the point of view of the consumers, because in the end, they are supposed to buy the product. Through this special interdependent relationship, every party in a vertical agreement can, in principle, be considered a natural ally of the consumer.

It is useful to approach this key observation in terms of an analogy: the difference between substitute products and complementary products (cf. Baxter, 1990). Whereas horizontal agreements concern agreements concluded by companies providing competing, substitutable goods or services, vertical agreements are concluded by companies that are providing complementary goods or services[8].

Where goods and services are *substitutes*, companies providing these goods or services are in direct competition with each other. When one firm cuts its price this will have a negative effect on the profits of other firms in the market as the demand for the products of those firms will fall. This effect is an external effect, in the sense that the price cutting company will, normally speaking, not take it into account. Each firm in the market has an interest in seeing the prices of the substitute products being increased. A joint profit maximizing agreement between the two (a cartel) will then seek to internalize the price externalities and lead to a joint increase in the prices. It goes without saying that the customers of these companies are hurt by such an agreement.

When goods and services are *complements*, the effects of price cuts are quite the opposite. When one company cuts its price this will tend to have a positive effect on the profits of suppliers of the complement by stimulating demand for their products. This effect is again an external effect, in the sense that the price cutting company will, normally speaking, not take it into account. Now, each firm has an interest in seeing

[8]Such a complementarity is most apparent in the interdependence between production and distribution: when a producer produces a good the only way the product can reach the market is through some form of distribution. On the other hand, each distributor is dependent on supply of products from the part of the producers.

the prices of the suppliers of complementary products being reduced. A joint profit maximizing agreement between the two firms will then seek to internalize the price externalities and lead to a reduction of the prices. This is exactly in the interest of the consumers. As a result, an agreement entered into by providers of complementary products is unlikely to be bad for welfare[9].

During the 1980s and 1990s, the Chicago methodology of studying the rationale of observed behaviour on the basis of rigorous micro-economic analysis (and its emphasis on the use of an economic efficiency criterion to evaluate the impact of such behaviour) has gone to the centre ground of Industrial Organization[10]. Its sharpest assumptions and conclusions, however, have not. For a large part this can be attributed to the increasing role of non-cooperative game theory in industrial organization, which allowed for the construction of a wider range of models and a wider range of outcomes. These models suggested that vertical integration or contractual restraints could be both rational and effective ways to engage in anti-competitive behaviour. For example, it was shown that delegating pricing decisions to exclusive distributors might allow producers to credibly commit to less competitive behaviour towards each other (to 'soften' competition), making use of the fact that the incentives to compete on the distribution level differ from those on the producer level[11]. Similarly, some exclusive dealing contracts were shown to be possible tools for foreclosing markets, in particular because they render the anti-competitive objective (i.e. foreclosure) more credible

[9]The importance of the products being complements rather than substitutes can also be more formally illustrated, along the lines of Baxter (1990). Let p_1 and p_2 be the prices of the two complementary products, π_1 and π_2 the profit levels of companies 1 and 2, respectively, CS_1 and CS_2 the consumer surplus associated with products 1 and 2, respectively and SW social welfare, the sum of consumer surplus and profits. The derivative of social welfare with respect to the prices of each of the two goods can be written as follows:

$$\frac{\partial SW}{\partial p_i} = \frac{\partial \pi_1}{\partial p_i} + \frac{\partial \pi_2}{\partial pi} + \frac{\partial CS_1}{\partial p_i} + \frac{\partial CS_2}{\partial p_i}$$

where $i = 1, 2$. What this expresses is that the change in social welfare when p_1 (resp. p_2) varies is the sum of the change in profits and consumer surplus associated with both products. When the two firms are setting prices non-cooperatively it follows that they will price according to the first order conditions $\partial \pi_1 / \partial p_1 = 0$ and $\partial \pi_2 / \partial p_2 = 0$, respectively. The terms $\partial CS_1 / \partial p_1$ and $\partial CS_2 / \partial p_2$ are negative: price rises are obviously negative for consumer surplus. Furthermore, when the two products are complements, the cross derivatives $\partial \pi_2 / \partial p_1$ and $\partial \pi_1 / \partial p_2$ are negative. Using these elements, it follows that at, at the non-cooperative equilibrium, the signs of $\partial SW / \partial p_1$ and $\partial SW / \partial p_2$ are also negative. This means that, at the non-cooperative equilibrium, each firm has an incentive to influence the price of the other into a direction that improves both consumer surplus and welfare: the private and the social incentives are aligned.

[10]Cf. Van den Bergh (1997).

[11]Cf. Rey and Stiglitz (1988, 1995).

and time-consistent[12].

As a result of the above development a consensus has arisen that it is impossible to predict the competitive and welfare effects of vertical restraints out of the context in which they are applied; there are circumstances in which they improve the efficiency of supplier-distributor relationships and increase competition, but there are also circumstances in which they may indeed be anti-competitive. Consequently, when it comes to vertical restraints, the move has been away from the traditional 'Chicago view' to advocacy of a more explicit balancing test, based on the circumstances of each case. As Tirole (1988) puts it 'theoretically, the only defensible position on vertical restraints seems to be the rule of reason. Most vertical restraints can increase or decrease welfare, depending on the environment. Legality or illegality per se thus seems unwarranted'. According to Kay (1990, p.560), 'the best conclusion is that we should principally look at the consequences, rather than the form or first order effects of the restraints'.

While this approach makes much sense from the viewpoint of economic theory, it has been recognized - not least by the cited authors themselves - that it would put far too heavy a burden on the antitrust authorities. An investigation into the effects of all the agreements that are concluded between firms at different levels of the production or distribution chain is just impracticable. Guidance in the form of a relatively robust characterisation of the circumstances in which vertical restraints are likely to have detrimental effects is therefore necessary to allow for antitrust supervision that is not only effective, but also efficient in keeping down administrative costs.

Seabright (1998) also points to a different reason as to why it is important to develop such guidance: to limit the costs that go with entrusting government authorities with discretionary powers. Seabright refers, among others, to the following 'costs of discretion', which are all fairly familar from the field of public economics. In the first place, discretionary policy may make it hard for firms to predict what the authorities are going to decide in particular circumstances. From the viewpoint of legal certainty, essential to any market economy, this is not desirable. Furthermore, the discretionary powers may incite both the authorities and the companies to engage in rent-seeking behaviour (lobbying on the part of the companies; cultivating 'friendly' relations on the part of government officials or governing political parties). It is well documented that often the most effective lobbying comes from the side where the economic stake

[12] Cf. Aghion and Bolton (1987).

is bundled in the hands of a few (industry), rather than from the side where interests are dispersed among many (consumers). Finally, even with a competition authority having the best intentions, the exercise of discretion bears the risk of inconsistency between decisions and between different branches of competition policy, with consequent distortionary incentives for the way in which firms plan their business strategies (see Chapter 3 for some examples in Europe).

As to the characterisation of the circumstances in which vertical restraints are likely to have detrimental effects for competition and welfare, the economic literature offers very extensive material on which to base such a characterisation[13]. The purpose of this chapter is to provide an overview of the main insights that economic theory has provided. Rather than giving a long enumeration of the different market situations that have been studied and the corresponding results, the aim is to present and develop the main arguments, occasionally with the use of some simple examples and models. In this, we will link up with the two main groups of motives for vertical restraints touched upon above: the vertical co-ordination motives (Section 2.2) and the anti-competitive motives (Sections 2.3 and 2.4, on foreclosure and 'softening' competition, respectively)[14]. A conclusion will follow in Section 2.5.

2.2 The vertical co-ordination motives

The simple analysis that identifies an essential difference between vertical and horizontal agreements (cf. Baxter, 1990) bears most of the argument that pleads in favour of vertical agreements in the economic literature: vertical agreements serve to

[13]The literature on vertical restraints has largely been framed in terms of the search for vertical control within a principal–agent relationship, where the principal (the supplier) imposes contractual obligations on its agent (the distributor) when delegating responsibility for selling its products. However, in contrast with the literature on mechanism design which seeks to characterise the outcomes that a supplier-distributor pair can achieve using general contracts, the literature on vertical restraints has traditionally taken contract obligations that are observed in practice as a starting point (exclusive distribution agreements, exclusive dealing, resale price maintenance, etc.). A good overview of general contract theory is provided by Salanié (1997).

[14]The classification into 'vertical co-ordination motives' and 'anti-competitive motives' (which is, in fact, common in the literature) might suggest that the former motives are not 'anti-competitive'. In the sense that the motives are intended to reduce inefficiencies within the vertical structure, rather than to hurt rivals outside the structure, this might be a proper interpretation. However, as will transpire from the analysis in section 2.2, it is not to be taken for granted that vertical restraints taken in pursuit of 'vertical co-ordination' will be beneficial for welfare. This may particularly be the case when 'vertical co-ordination' also allows for a better exploitation of monopoly power. In those cases, the vertical restraints may just as well be deemed 'anti-competitive'.

co-ordinate the actions of an upstream firm and a downstream firm and they may well be welfare inproving in view of the complementary nature of the relationship. At the same time, it does not capture the complexity of most market situations and a more developed analysis is required to account for this complexity. For example, what are the welfare effects of the respective vertical restraints in settings in which companies take decisions on more dimensions than just the price dimension (e.g. service levels, advertisement) ? And what if there are more players on either level ? What are effective ways to co-ordinate when it is difficult or even impossible to write contracts that take into account every possible contingency ? In this section, these variations will be discussed in turn.

At first sight, it would seem that a supplier (manufacturer) would be keen on its distributors (retailers) being as competitive as possible: all other things being equal, the smaller the distributor mark-up, the greater the sales and profit levels for the supplier. However, this is very much like wishful thinking. Distributors incur fixed costs, many of which contain a sunk element. The fact that there are fixed costs means that perfect competition is an unattainable ideal framework for the distribution level. As a result, suppliers will normally be faced with a distribution level which has some market power (in the sense that they can set prices above marginal cost) and which is, by consequence, in a position to choose actions which are not in line with the interests of the consumers or of the supplier. The distributor's possibility to choose actions 'for itself' may lead to externalities of the type illustrated by the analysis of Baxter (1990).

2.2.1 Controlling the basic vertical externalities

Let us, in first instance, abstract from the interaction with other suppliers and distributors and focus on the possible co-ordination problems within a structure which is made up of one supplier (upstream) and one distributor (downstream)[15]. In its simplest setting, with consumer demand only depending on the retail price, we obtain the classic problem of double marginalisation (Spengler, 1950). Let $D(p)$ denote market demand as a function of the retail price. Let us suppose that the marginal cost of production is c and that, for simplicity, the distributor incurs no cost other than the wholesale price p_w it has to pay to the supplier. Suppose further that the manufacturer supplies the distributor at a constant wholesale price and that the distributor

[15]The structure of the analysis provided in this subsection is along the lines of Tirole (1988).

can determine the retail price independently. Then, for a given wholesale price, the distributor will charge the retail price that maximizes its profit $(p - p_w)D(p)$, i.e. it will charge the corresponding monopoly price $p^m(.)$, which is a function of the wholesale price p_w. To make a profit, the supplier will charge a wholesale price that exceeds its marginal cost of production: $p_w > c$. However, because of the two successive margins (both $p > p_w$ and $p_w > c$), the retail price ends up too high from the viewpoint of the structure as a whole: the retail price is $p^m(p_w)$, with p_w chosen optimal by the supplier, whereas it should optimally be $p^m(c)$, as c is the 'true' marginal cost of the vertical structure. The pricing distortion arises from the fact that the distributor does not take into account the effect on the profit stream flowing to the supplier; nor does the supplier take into account the profit stream flowing to the distributor, for that matter.

Vertical integration in the usual sense of the expression, namely common ownership of both firms, would internalize this effect, but alternative contractual relationships also solve the problem. Using Tirole's (1988) terminology, the 'target' of the vertical structure is to fix the retail price at the right level. One way of doing so would be for the manufacturer to use resale price maintenance and fix the price at $p = p^m(c)$. A condition to use this instrument is that the retail price is observable by the supplier and verifiable, i.e. suitable to be written down in a contract. Another way would to avoid the pricing distortion would be to use a two-part tariff, consisting of a marginal wholesale price equal to marginal cost (c) and a fixed fee that recovers the distributor's subsequent operating profit. The distributor is then made the residual claimant of the vertically integrated profit: it captures the entire benefit of every extra unit of product sold. A final way to solve the problem would be to sell at a wholesale price equal to the target price and impose quantity forcing. All three steps lead to a lower retail price, an increase in profits and an increase in consumer surplus.

A variation of the double marginalisation problem discussed above results when 'promotional effort' or 'services' provided by the distributor enhance the value of the product to the final consumer[16]. In these circumstances, demand will depend on both the price level p and on s, the level of service: $D = D(p, s)$. Suppose that it costs the distributor an amount $\gamma(s)$ per unit of output to offer these services. A vertically integrated structure would choose the price-service combination (p^m, s^m) that maximizes the integrated profit $(p - c - \gamma(s))D(p, s)$. A distributor however

[16] A similar argument can be built when the supplier takes care of promotional effort.

maximizes only its own profit, $(p - p_w - \gamma(s))D(p, s)$. Whenever the supplier charges a wholesale price that exceeds the true marginal cost of production, there will be a distortion similar to the one observed in the simple example of double marginalisation: both in choosing the level of service and in the level of the retail price, the distributor will not take into account the profits flowing to the supplier. In particular, as the distributor does not reap all the benefits from increasing the level of service, it will typically choose a level that is too low from the viewpoint of the structure. A supplier should therefore not opt for regular market transactions at a certain unit price, but opt for some form of control.

If s or $\gamma(s)$ were observable and verifiable, a contract could be written that explicitly specifies the appropriate amount of service, s^m, to be provided. In combination with retail price maintenance at a price equal to p^m, the vertically integrated profit could be realized. A different, less involved way, would again be to make the distributor a residual claimant by selling him the input at marginal cost, $p_w = c$, and to appropriate the distributor's operating profit through a fixed fee equal to the vertically integrated structure's profit. An alternative way to ensure that the distributor provides the appropriate level of services would be to write a quantity forcing contract at $D(p^m)$.

The welfare effects of the possible vertical restraints in this setting are ambiguous (cf. Comanor, 1985). Whereas the consumers will appreciate the internalization of the double marginalisation effect (which leads, all other things being equal, to a decrease in the retail price), the internalization of the service externality leads, all other things being equal, to an increase in the service level. The resulting increase in the service costs may lead to a net increase in the retail price. Depending on the exact specification of the demand structure, it cannot be excluded that the situation without vertical control is the preferred one by the consumers. The underlying reason is that the exercise of market power creates a division of interest between the marginal customers (customers that are at the brink of not buying the product) and infra-marginal customers (ones that are buying the product even if the price rises a little or the level of service goes down a little). In search of an increase of profits at the margin, the vertically integrated structure tries to make extra consumers buy the product, e.g. by further increasing the level of services, as long as this is profitable. However, whereas both marginal and infra-marginal customers share an interest in the provision of a given product and service at lowest cost, they are likely to have divergent preferences regarding the exact combination of price and service (this is

precisely why some consumers are at the margin and others are infra-marginal). For instance, the infra-marginal consumers who are buying the product anyway, may rather prefer to pay lower prices and spend the money saved on something else. As a result, is not necessarily the case that a vertical agreement leading to an increase in service provision will be welfare enhancing. The simple framework of Baxter (1990) does not easily translate towards situations with more decision variables than just the price level.

2.2.2 Controlling externalities between distributors

In the above analysis, we saw that vertical co-ordination is useful when one of the two parties, say, the distributor, does not reap all the benefits from its actions, whereas it does bear the costs. It was the upstream firm in the vertical structure who was the co-beneficiary of the 'proper' actions taken. When there are more companies active in the distribution of the supplier's product, another set of externalities arises, externalities between distributors.

Whenever distributors raise the price at which they sell, they confer benefits on competing distributors of the product, as more consumers will turn to these other distributors. An individual distributor will normally not take this externality into account and hence set prices lower than would be optimal for the distribution level. Whereas this is generally a good thing for the supplier and the consumers alike, some inefficiencies may result. One possible concern is the number of distributors willing to carry the product[17]. When there is differentiation between distributors, for example as a result of their location or their marketing format, aggregate demand for the product will increase in the number of distributors (all other things being equal). When price competition between distributors is too strong, this may take away the incentive to set up a distribution business alltogether. This is, all other things being equal, not in the interest of the supplier. The supplier might then want to use price restraints (resale price maintenance) in order to support the number of sales outlets. Alternatively, the supplier could use fixed subsidies (e.g. slotting allowances) to influence the number of distributors carrying its product. The exact trade-off between the number of sales outlets and the use of resale price maintenance has been studied in a number of settings, but the theoretical support for the argument that the number of distributors

[17] Cf. Yamey (1954), Dixit (1983), Matthewson and Winter (1983), Bittlingmayer (1983), Ippolito and Overstreet (1996).

ends up too low because of price competition is, in fact, limited (see Chapter 4 for a more general discussion). For an argument of this type to hold, richer structures appear to be necessary.

Externalities among distributors also occur in the domain of services provision and promotional effort[18]. Product specific advertising by one distributor may well benefit other distributors of the product. The same holds when customers can obtain in-store information from one distributor and buy the product from another. To the extent that some of the benefits of service provision are enjoyed by other distributors, distributors will want to free ride on each other and provide less service themselves than would otherwise be the case. This effect is reinforced by the fact that such behaviour allows for more aggressive pricing from the part of the free-riding firm as it does not incur the costs related to services provision (e.g. a discounter)[19].

To encourage an adequate provision of services and promotional effort by distributors several options are available to the supplier[20]. It can choose to eliminate price competition at the distributor level by imposing resale price maintenance. Similarly, exclusive territories are a means to reduce the externality problem. More generally, in fact, any measure that reduces the intensity of price competition between distributors is a way to reduce the problem (e.g. with differentiated distributors: operating a selective distribution system, charting sufficiently high fixed fees).

While the above measures will reduce the level of free-riding behaviour by reducing the scope for price cutting behaviour, they need not entirely take away the externality. In the case of advertising, for example, there is still the problem that not all of the benefits of are advertising are reaped by the company engaging in it. In this case, it may be necessary to provide additional incentives to advertise, e.g. by charging a very low wholesale price (possibly even below marginal cost, while appropriating the distributor's operating profit by means of a fixed upfront payment). In the case of in-store information provision, resale price maintenance and exclusive distribution may, however, already be sufficient as it is unlikely that customers, having obtained the

[18] Cf. Telser (1960), Mathewson and Winter (1984), Besanko and Perry (1997), among others.

[19] The effects are most easily seen from the case of two distributors and product advertising. Suppose market demand is given by $D(p, s)$, where s is the sum of the expenses made by the two firms for advertising ($s = s_1 + s_2$). Suppose also that consumers buy from the distributor which charges the lowest price. In that case, there is no equilibrium in which one of the two firms will advertise. The only equilibrium will be one in which no advertisement is provided at all.

[20] We will focus on services and promotional effort which are best provided by distributors, such as in-store information provision or location-specific advertising. After all, if the supplier could equally well take care of promotional effort himself, he could simply choose to do so.

pre-sale information from one distributor will then go to another if this one charges the same price or if it is located far away.

As for the welfare effects of the restraints that are used to internalize the service externalities between distributors: if the divergence of interest between marginal and inframarginal consumers is not too large, it is likely that the restraints are welfare enhancing, as consumers will also appreciate the resulting increase in the level of services.

2.2.3 Avoiding externalities that benefit other suppliers

The previous section highlights the role of vertical restraints to control for externalities among distributors. By just considering one supplier and its distributors, they abstract from concerns that the supplier may have about the externalities that are at play at the upstream level.

Important from the perspective of internal efficiency are the externalities that go with efforts from the part of the supplier to improve the distribution channel when also other suppliers make use of this channel for the distribution of (part of) their sales[21]. For example, when a supplier provides commercial or technical training to its distributors, this makes these distributors not only more effective in selling the supplier's product but also at selling its rivals' products. The training, therefore, also works to the rival suppliers' benefit. In addition, because these rivals do not incur the costs of training, they are in a position to charge lower wholesale prices and obtain extra market share. As a consequence of this free-rider problem, suppliers can be expected to make investments that are too low from the viewpoint of the vertical structure.

In order to eliminate this externality, the supplier may want to use exclusive dealing contracts, as a result of which the distributors cannot deal in the competitors' products. While it may be true that the supplier still does not capture all the benefits of its efforts (the distributor may still use the knowledge for other purposes), the benefits no longer accrue to its direct competitors. As such, exclusive dealing is a means of providing the supplier with a kind of property right to the fruits of his efforts.

Marvel (1982) notes that the upstream externality problem is likely to arise in situations where the opinion and recommendation of the distributor carries a certain

[21] Cf. Marvel (1982), among others.

weight in the consumers' purchase decisions (otherwise, the supplier will not bother to provide training in the first place). It is precisely in such a setting that the downstream service externality can play a role. According to Marvel, this might form an important explanation for the fact that exclusive distribution agreements and exclusive dealing agreements often go hand in hand. Furthermore, he notes that franchise agreemeents are often accompanied by non-compete obligations, as they involve the transfer of a substantial amount of know-how generated by the supplier.

A different efficiency motive that has been given for exclusive dealing arrangements is that they ensure that the distributors market the products 'with maximum energy and enthusiasm'[22]. If a distributor markets several products it will spend its efforts in such a way that the marginal revenue of effort will be equal acrosss the products. A manufacturer may then choose exclusive dealing in order to elicit distributor effort for its product. This argumentation does have a weakness, however. After all, a supplier might also choose to reduce the wholesale price in order to elicit more effort (possibly accompanied by a fixed fee).

2.2.4 Vertical co-ordination and risk sharing

As can be seen from the above sections, vertical restraints can be used to align incentives in cases where these are not aligned. One reason for the divergence of interest are the respective horizontal and vertical externalities. In the above sections, these externalities occurred in situations where both the supplier and the distributors are similar in the sense that they are all profit-maximizers. Another reason why there may be divergences of interest is that the supplier and the distributors may have differing attitudes towards risk. In that case, the two levels of the vertical structures make different trade-offs between profit maximization and the risk they want to bear.

It turns out that while different types of vertical restraints may bring about similar results in deterministic environments (see the above sections[23]), this is less the case in risk environments. For risk considerations to play a role, one should not only presuppose that there is uncertainty about the level of some parameter but also that there is an asymmetry of information about this parameter. After all, if a supplier could

[22] Scherer (1980, p. 586).

[23] This is not to say that in deterministic environments vertical restraints always have similar effects. Cf. Marvel and McCafferty (1996) for resale price maintenance vs. exclusive territories or Reiffen (1999) for resale price maintenance vs. quantity ceilings.

observe, ex-post, the realisation of all parameters relevant to the distribution stage, he could structure the incentives for the distributor as in a deterministic environment, using contracts which are conditional on these parameters[24].

In order to explore risk considerations in the context of vertical restraints, Rey and Tirole (1986) set up a spatial model of retail competition and analyse the role of the restraints when there is uncertainty about future demand and cost levels. The retailers are better informed (ex post) about the realisation of final demand and about their own costs than the supplier. The latter is assumed not to be able to obtain this information, directly or indirectly (one of the reasons for this is that the retailers can engage in arbitrage, making it impossible for the supplier to keep track of the supplies). Hence, informational problems prevent the supplier from using contracts based on the true performances (profits) of the distributors.

Rey and Tirole consider two vertical restraints, resale price maintenance and exclusive territories, and compare them with the situation of competition between differentiated retailers. The basic trade-off in the choice of contract is between the optimal exploitation of market power and the amount of risk that the distributors are willing to accept. The optimal exploitation of market power requires that one makes optimal use of local information and that one avoids the double marginalisation problem associated with linear wholesale pricing. With uncertainty, the supplier is best able to exploit its market power by providing exclusive territories since the retailers, as local monopolists (when they are made residual claimants for the profits of the vertical structure), adjust their prices optimally to cost and demand conditions. If instead the supplier imposes resale price maintenance, the retail price is fixed and, hence, cannot repond to cost and demand shocks. Without either of these restraints, competition ensures that the retail price is principally driven by the cost level, rather than by the demand conditions, even though the latter also have some influence.

Whereas the ranking in terms of exploitation of market power is rather clear, the risk properties are more complex. When the retailers are more risk averse than the manufacturer, internal efficiency requires that the supplier shares some of the risk with the distributors. Competition provides the distributors generally good insurance: as the retail price is mainly cost driven, the response to cost shocks is good. At the same time, as their margins are anyway not very large, the variability of their profits to

[24] In fact, this argument holds if the relevant parameters are observable and verifiable, i.e. suitable for use in a contract.

demand conditions is moderate. The insurance properties of resale price maintenance depend on the type of uncertainty. With demand uncertainty, resale price maintenance may provide rather good insurance, in particular if the distributor margin is kept low by the supplier. With market wide cost shocks, distributors bear the whole risk as they cannot adjust the price level to these shocks. On the other hand, with distributor-specific cost shocks, the picture is mixed: whereas resale price maintenance remains a rigid instrument, it may provide insurance in the sense that they protect distributors against more efficient distributors. Exclusive territories generally have poor insurance properties, despite the fact that exclusive distributors can adjust their prices to meet cost and demand changes: as their incentives are structured to be those of residual claimants, they are faced with substantial fluctuations in their profits.

The specific choice of vertical restraints depends on the trade-off between the insurance objective and the optimal exploitation of monopoly power. When risk is not too important, exclusive territories are likely to an optimal way to adjust to local circumstances. With very risk averse distributors, the insurance objective dominates the desire for optimal exploitation of market power. The nature of the market uncertainty, therefore, determines the supplier's choice whether or not to use vertical restraints at all and the ranking of the preferred restraints.

2.2.5 Reducing transaction costs

¿From one point of view, the theory of vertical integration and vertical restraints is just a special case of the theory of the firm. An alternative way to look at vertical linkages is that they are a way to minimize transaction costs, or to reduce them below market transaction levels[25]. Transaction costs can be understood as the usual costs of searching a trading partner and of drawing up and enforcing contracts, but also as inefficiencies that result from not being able to write contracts as comprehensive as one might wish (incomplete contracts).

In the first interpretation, the costs that result from searching a trading partner may lead companies to engage in longer term contracts, whereas the desire to reduce the costs of drawing up and enforcing contracts may lead firms to limit the number of trading partners. Similarly, the logistics of dealing with trading partners may lead to a limitation of the number of trading partners, also in time. Frequency of trade is

[25] Cf. Coase (1937), Williamson (e.g. 1975, 1989). Formalized by Grossman and Hart (1986); refined by Holmström and Tirole (1988).

also an important element: if the same sort of bargain is to be concluded quite often, then it may make sense to devise a contracting framework to facilitate it.

The main elements in the second interpretation of transaction costs are the concepts of contract incompleteness and asset specifity. A contract is called *complete* when it covers all possible contingencies that may occur and has the relevant decisions (on price, quantity, product characteristics, etc.) depend on verifiable variables (including, possibly, announcements by the parties, e.g. concerning their valuations, costs, etc.). On the other hand, when the contract does not cover all possible contingencies and, hence, leaves open what to do in those cases, a contract is incomplete. At the origin of contract incompleteness is, of course, that it may be very costly (or even impossible) to write complete contracts.

Assets are specific to a transaction to the extent that they are more valuable within the scope of the transaction than outside the scope of the transaction. An example of asset specifity is when a distributor makes investments (e.g. in retail format) to suit the product range of the supplier and these investments are of little use for the distribution of other suppliers' products. Similarly, the provision of technical and commercial training provided by a supplier to a distributor is most useful when this distributor is effectively going to distribute for the supplier and not for some other supplier.

A crucial aspect of specific investment is that even though the supplier and the distributor may select each other ex ante from a pool of competitive suppliers and distributors, they end up forming an ex post bilateral monopoly in the sense that they have an incentive to trade with each other rather than with outside parties. However, each party also knows that, if there are no checks and balances on each other's behaviour, the parties may have an incentive to enter into opportunistic behaviour the moment at which the other party has invested, in an attempt to obtain a greater part of the surplus that is created: to the extent that a substantial part of the value of the investment has become 'stuck' in the relationship, the party that has invested a lot finds itself in a weak bargaining position vis-a-vis the party that has not invested (as much) in the relationship. This perspective, the weak bargaining situation ex post, is likely to change the incentives to invest ex ante and to lead to investment levels that are too low from the viewpoint of the vertical structure.

In order to avoid these inefficiencies, it may be optimal for the parties to enter into contracts ex ante. If complete contracts could be written, the solution would be trivial: specify in the contract the investments that should be undertaken and the

terms and conditions under which trade is to occur. In fact, it would not even be necessary to say anything about the investment levels in the contract. In particular, pre-specified selling prices for the products of the investing supplier (whenever there are effectively gains from trade), will do; in this case, the supplier is ensured that the distributor will not engage in opportunistic behaviour[26]. Similarly, when it is the distributor who must make the specific investments, a guaranteed purchase price (whenever there are effectively gains from trade) will be sufficient.

However, when it is too costly or simply impossible to write complete contracts, internal efficiency cannot be attained (save in a few exceptional circumstances); after all, in the cases that a non-specified contingency occurs and ex post bargaining must occur, the party that has less at stake may engage in the types of opportunistic behaviour described above. In the realm of vertical restraints, one way to reduce the incentives to engage in such behaviour would be to enter into contracts that provide for some kind of exclusivity. For example, when the distributor is the party that has to make the specific investments, a contract granting an exclusive territory to this distributor has the effect of reducing the outside opportunities for the supplier. As a result, the bargaining position of the distributor is improved and its incentives to invest have increased. Similarly, when the supplier is the party that has to make specific investments, an exclusive dealing contract prohibiting the distributor to deal in rival products may have the effect of reducing the outside opportunities for the distributor.

Despite the insights obtained from the transaction cost / incomplete contracting approach, it is unfortunate that the approach typically does not address the question whether there are possible divergences between the private and social desirability of vertical restraints (the literature typically focusses on the private incentives of firms). Nonetheless, according to Williamson, the transaction cost approach shows that vertical links can yield cost savings over a wider range of circumstances than the earlier market power approach indicated and that a more permissive view of vertical integration is warranted.

[26] One may note that the transaction cost literature, with its emphasis on restoring the ex-ante incentives to invest, has a flavour similar to the externalities-based literature. In a way, it is just the source of the investment problem which is different.

2.3 The anti-competitive motives (I): foreclosure

The various settings considered up to this point address the ways in which vertical structures seek to improve the internal efficiency of the structure. While there are exceptions, it transpires that, generally, vertical restraints inspired by an internal efficiency motive may well improve welfare. There exists, however, also a body of literature that has considered the ways in which vertical structures attempt to reduce the competition with rivals and the circumstances in which this is profitable.

This literature can be divided into two main lines of thought. The first is that vertical restraints may lead to foreclosure of market access, leading to a reduction of the ability of other firms to compete. The second line of thought is that vertical restraints may be used to 'soften' the competition between suppliers or even to enact and enforce outright cartels.

2.3.1 Foreclosing market access to rival suppliers

Exclusive contracts have long been considered in the literature as practices to reduce market access to actual or potential competitors, thereby increasing their costs. This motive is interchangeably referred to as 'foreclosure', 'exclusionary behaviour' and 'raising rival's costs'. Most attention in this field has focused on the role of exclusive dealing contracts, contracts that prohibit a distributor to deal with other suppliers.

A supplier's profits are typically an increasing function of its rivals' costs and prices. When there are substantial economies of scale and scope in distribution, signing exclusivity contracts with particular distributors raises the distribution costs of other suppliers and reduces the possibilities to reach the market for new suppliers (entry barriers). In the extreme, when there is only one distributor available, exclusive dealing arrangements have the effect of completely foreclosing the market.

Whereas the above effects of exclusive restraints appear rather straightforward, the rationality of foreclosure is not to be taken for granted. Indeed, one question that has frequently been raised is the following: why would a distributor, in particular when he is a monopolist, be willing to enter into exclusive dealing agreements with suppliers and forego the opportunity to deal (also) with other, possibly more interesting suppliers? As Bork (1978) points out, the distributor would only accept such an arrangement if it were compensated in some way. In this sense, the problem may not be so much that the supplier will not desire to exclude a competitor, but that the

cost of inducing a distributor to sign an exclusivity arrangement may not make such a restraint profitable.

One formalization of this intuition has been provided by Mathewson and Winter (1987)[27]. In essence, they consider a model in which two suppliers sell differentiated products through a local monopoly distributor. Under exclusive dealing, the suppliers compete on the basis of wholesale prices for the right to be selected by the distributor. Complete foreclosure occurs when one supplier offers an exclusive dealing contract with a wholesale price low enough to make it more profitable for the distributor to accept this contract than to accept a contract from the other supplier at a wholesale price equal to marginal cost. In this model, exclusive dealing allows the efficient supplier to exercise its existing cost advantage in production to exclude the other supplier, but, depending on the demand characteristics, it will not always choose to do so.

In fact, the issue of the rationality of foreclosure touches upon a more general point. As explained in the introduction to this chapter, competition between companies operating at the same level entails a horizontal externality: all companies would be better off if they would behave less competitively. An obvious way to internalise this externality is to vest all of the decision-making power in a single economic actor. So, why does foreclosure occur in the first place, if the distributor may as well serve as a common agent ?

Bernheim and Whinston (1985, 1998) have formally explored this issue in a setting similar to, but more general than that of Mathewson and Winter (1987). Consider two suppliers of differentiated products that both bid for representation by a single distributor. Each bid consists of two parts: a wholesale price schedule in the event the supplier is represented exclusively and a wholesale price schedule in the event the distributor represents both suppliers. A priori, no restrictions are placed on the types of price schedules that can be used[28]. In such circumstances, Bernheim and Whinston show that despite the fact that the suppliers are acting non-cooperatively, the form of representation will result that maximizes the joint surplus of the suppliers and the distributor achievable under the circumstances: as each supplier must effectively compensate the distributor to attract it to an exclusive deal, suppliers internalize the distributor's cost from the loss in product variety.

[27] Another, slightly less sophisticated formalization is that of Comanor and Frech (1985). The contribution of Mathewson and Winter (1987) was a response to their article.

[28] Mathewson and Winter (1987) only considered linear wholesale tariffs.

Having established his principle, Bernheim and Whinston show that the chosen form of representation critically depends on the existence of contracting inefficiencies resulting from the non-cooperative provision of incentives by the two suppliers under common agency. By way of definition, contracting inefficiencies are present when the outcome under common agency is less than the outcome that would arise if the suppliers were to cooperate in their contract offerings. When there are no contracting inefficiencies, common agency will necessarily be the optimal distribution form: after all, selling both products is jointly more profitable than only selling one (whenever products are differentiated). In such circumstances, the joint surplus can be obtained by delegating the actions to the agent and making him the residual claimant. This entails an optimal wholesale price equal to marginal cost and a franchise fee set to drive the agent's net profits to zero. In essence, the business is effectively 'sold out' to the agent[29].

It is only when there are certain contracting inefficiencies that exclusive dealing may arise in equilibrium. One such contracting inefficiency underlies, in fact, the result of Mathewson and Winter (1987). In their setting, each supplier is restricted in its choice of contract, in the sense that it can influence the distributor only by means of a linear wholesale price. This implies that the suppliers cannot use the type of 'sell out' contract described above: for a supplier to make some profit, it is necessary to keep a positive wholesale price margin. Because a positive wholesale margin distorts the pricing incentives of the common agent, the agent will not maximize industry profits and will not perceive the full opportunity costs of exclusive dealing (in equilibrium, the agent's perceived opportunity cost of exclusive dealing is less than the corresponding value to the suppliers)[30]. Other circumstances in which contracting inefficiencies can arise is when incentive provision is costly because of informational asymmetries that lead to adverse selection or moral hazard problems[31]. Finally, Bernheim and Whinston

[29]Note that in the corresponding formal game there is a continuum of subgame perfect equilibria; the equilibria only differ in the division of profits between the two suppliers.

[30]The restriction on the feasible set of contracts in Mathewson and Winter (1987) has another, noticeable effect: even though the two suppliers are competing for the right to be selected, the distributor ends up worse off than in the case of unrestricted dealing: inducing a distributor to sign an exclusivity arrangement does not necessarily imply full compensation of this distributor. In this sense, Mathewson and Winter only provide for a partial formalisation of Bork's setting. Another difference is that they assume that the distributor is one of the many local monopolies that are being served by the 'dominant' supplier and that, as a result, the relative bargaining power of the distributor is negligible. This implies that the distributor is not able to 'force' the two suppliers to deal with him.

[31]See also Martimort (1996).

identify another possible source of contracting inefficiencies: when suppliers must serve more than one market to achieve scale economies, exclusion may result because the signing of an exclusive contract in one market has knock-on effects that are not fully being taken into account by the distributor.

In the above settings, exclusive dealing will not occur unless there are some contracting inefficiencies arising from two suppliers simultaneously bidding for representation. It does appear, however, that a different perspective arises when one supplier, for instance an incumbent supplier, has a first-mover advantage. For instance, Aghion and Bolton (1987) consider a monopoly supplier who faces potential entry into its market. Whether or not entry will occur is uncertain: the efficiency level of the entrant is not known to the incumbent supplier (or distributor) ex ante. Aghion and Bolton show that a long-term exclusivity contract between the incumbent supplier and the distributor that stipulates a penalty clause in the case of prior termination, is a profitable entry-deterrent. The optimal penalty clause has the effect of deterring entry, but not always: when the entrant's cost is very low, it will enter and reimburse the distributor for the penalty it has to pay to the incumbent supplier to break the contract. As a result, the contract has the effect of extracting (part of) the surplus of the future entrant. This expected benefit of the penalty clause allows the supplier to compensate the dealer in the form of a lower wholesale price for the goods to be traded. Note that, in this setting, it is not in the interest of the supplier to completely foreclose the market and always deter entry: the higher the penalty, the lower the risk of entry, but also the lower the expected payment to be received (at some point). There is, hence, a trade-off.

Comanor and Rey (1998) provide for another reason as to why a distributor may want to accept exclusive contracts in a dynamic framework. If, for example in the above described context, the incumbent supplier has the possibility of setting up a distribution network of its own (albeit, perhaps, at higher cost), the decision to switch to the entrant supplier may trigger competition between two vertical structures (the distributor and the entrant supplier vs. the incumbent supplier and its network)[32]. This competition has the effect of dissipating the profits in the industry. To the extent

[32]It is assumed that the incumbent supplier can distribute its product through a 'new' network, whereas the entrant supplier cannot. One interpretation is that the entrant supplier's product lacks market recognition and can only be succesfully sold through an established, influential distributor. The 'switch' to the entrant supplier may also be understood as the distributor also selling the entrant's product (in addition to the incumbent's product). The context of Comanor and Rey (1998) is focused on an incumbent distributor facing a low cost competitor, but the situation is comparable.

that the combined profit of the distributor and the incumbent supplier suffers, it is in their mutual interest to enter into an exclusive contract.

2.3.2 Foreclosing market access to rival distributors

It is also possible that distributors seek to conclude exclusivity contracts with a supplier in order to exclude rival distributors from the market. For example, when a supplier is the sole supplier in a product market, a contract that grants a distributor an exclusive sales territory has the effect of removing rival distributors as competitors. Similarly, it has been suggested by the literature that a distributor may raise its rivals' costs by signing up an exclusive contract with an upstream supplier when there is substantial market power at the upstream level[33]: by linking up with one supplier, the remaining suppliers end up having more market power and this translates into higher wholesale prices for the rival distributors. In this version, the theory of raising rivals' costs is one of creating double marginalisation. However, in the double marginalisation also lies its weakness: the result disappears as soon as the rival suppliers and distributors can use two-part tariffs to eliminate the double marginalisation problem.

As to the rationality of these arrangements, similar remarks can be made as in the context of exclusive dealing arrangements. For instance, if the distributors are differentiated (and consumers appreciate this), why would the supplier agree to deal with only one of them? The arguments of Aghion and Bolton (1987) and Comanor and Rey (1998) would appear to apply, mutatis mutandis, also to these situations: incumbent market players may want to avoid entry by new companies and a first-mover advantage may make it worthwhile to enter into exclusive contracts. By contrast, the arguments brought forward by Bernheim and Whinston (1998), which are based in essence on the observation that it is beneficial to avoid price competition by using a common agent, are less relevant in this context.

2.3.3 Foreclosure as a commitment device

Also when there is no need to raise rival's costs through scale economies, it is possible that a supplier may want to explicitly exclude some distributors from its market operations and deal with only one of them. This is a case of 'upstream foreclosure', in

[33] Cf. Ordover, Salop and Saloner (1990) or Krattenmaker and Salop (1986).

the sense that the left-out distributors no longer have access to the suppliers' products. According to Chicago economists such as Bork (1978), if a monopolist supplier were to use exclusive contracts in such a situation, this supplier would do it for efficiency reasons, not for the purpose of raising prices. After all, a monopoly rent can only be earned once: it is not possible to increase this rent through exclusive contracts. These arguments have been studied in detail by Rey and Tirole (1996), drawing on earlier arguments of Hart and Tirole (1990). Their main argument is that one should not take for granted that a supplier with market power is able to exploit this market power when it lacks the means to commit to do so[34].

Consider the case of a monopolist supplier M with just two distributors, D_1 and D_2. The distributors first order their quantities before they resell them on the market, which is characterised by Bertrand price competition. Clearly, the profit maximizing price and output for the vertical structure are given by the monopoly price P^m and quantity Q^m, respectively. Let us suppose that, for simplicity, distributors incur no costs other than the wholesale price to be paid to the supplier. The standard argument would then be that the supplier can set the wholesale price at the monopoly price P^m, so as to capture the entire monopoly profit: both distributors would agree to procure half of the monopoly quantity in order to sell it on the market. However, this argument does not work if there is scope for the supplier to offer different prices to different retailers and the wholesale contracts are secret (or secretly renegotiable). Intuitively, a commitment problem arises because, when dealing with one distributor, the supplier has incentives to free ride on the sales of the other distributor, as he does not internalize the latter's mark-up on those sales[35]. The distributor who is offered a wholesale price equal to the monopoly price anticipates this incentive and will not accept the offer[36]. This means that the supplier will not be able to charge

[34] O'Brien and Shaffer (1992) have studied the commitment problem using a less standard approach.

[35] To see this more formally, denote by $P(Q)$ the market clearing price at quantity Q. Suppose that D_1 expects that D_2 will purchase $\frac{1}{2}Q^m$ units from the supplier. Then D_1 will be prepared to pay a price of $P(q_1 + \frac{1}{2}Q^m) \cdot q_1$, for any amount of q_1, as this is the profit it expects to make when D_2 brings to market $\frac{1}{2}Q^m$. In that case, the supplier's profit maximization problem with respect to D_1 corresponds to maximising $P(q_1 + \frac{1}{2}Q^m) \cdot q_1$ over q_1. But this is exactly the way in which a quantity setting company would react to a given quantity set by its competitor. As the former does not take into account the profits of the latter, it will typically choose a quantity (*a price*) that is higher (*lower*) than optimal for the two firms seen a as a whole, i.e. a quantity higher than $\frac{1}{2}Q^m$ (*a price lower than* P^m).

[36] The assumption of 'passive beliefs' is important here, i.e. the assumption that distributors do not revise their beliefs when they observe out-of-equilibrium offers from the supplier. If the distributors had 'symmetric beliefs' (believing that the supplier offers identical contracts to all distributors), the

the monopoly price to its distributors and not realize the monopoly profit[37].

The above problem arises because the upstream supplier cannot commit to charge the same wholesale price to its distributors. A solution is then for the supplier to effectively 'tie his hands' by signing an exclusive distribution contract with one distributor. Industry-wide resale price maintenance, if this were enforceable, would be another option. Both vertical restraints would have the effect of taking away the incentive for the supplier to 'cheat' on the distributors and of restoring its market power. Obviously, the greater the market power to be restored, the greater the incentives to use vertical restraints as a commitment device.

2.4 The anti-competitive motives (II): softening competition

In the preceeding section we investigated the possibilities of firms to use vertical restraints with the objective to foreclose the upstream or downstream markets. The number of cases in which complete foreclosure is possible as a result of, for example, entry barriers may be limited. However, also in the absence of foreclosure possibilities, vertical restraints can be used to reduce competition in a strategic way. Two variants can be distinguished. First, vertical restraints may be used to 'soften' the competition between suppliers (section 2.4.1). Second, they may be used to enact and enforce outright cartels, at the supplier level or at the distribution level (section 2.4.2).

2.4.1 Softening competition

Because vertical restraints directly affect the nature of downstream intrabrand competition between distributors, they also affect the competitive behaviour of the supplier whose product is sold by the distributors. This, in turn, may alter the nature of upstream interbrand competition. In this sense, vertical restraints may be used strategically by a supplier as a means to commit itself to act in a certain way vis-à-vis its rivals.

commitment problem would disappear. As a variation, White (1999) considers the case in which distributors derive information on the (supposedly unknown) cost level of the rival distributor from the specific contract being offered.

[37] In fact, the supplier is only able to obtain the Cournot profit. It will end up maximising $P(q_1+q_2)\cdot q_1$ over q_1 and $P(q_1+q_2)\cdot q_2$ over q_2. The result is the same as that of a Cournot situation.

Rey and Stiglitz (1988, 1995) consider the case of two suppliers of differentiated products using exclusive territories (reducing competition among distributors carrying the same brand) as a device to reduce competition between themselves. When there is unrestricted competition on the distribution level for a particular product, the final price at which distributors sell closely follows the cost level faced by them and, in particular, the wholesale price set by the supplier. With exclusive territories, a distributor, as a consequence of being given some monopoly power, has more freedom in its price setting behaviour. When the supplier reduces its wholesale price, this distributor will pass on only part of the reduction to the customers. Furthermore, when distributors of the rival product also have some power over price and can observe the reduction of the wholesale price, these distributors will react by lowering their selling price. Both effects lower the increase in demand that can be expected by a supplier when it reduces its wholesale price, thus discouraging such a reduction in the first place. The rival vertical structure will anticipate this and behave less competitively in selling its product. As a result, the final selling prices and profits will end up higher than would be the case if the products were distributed through unrestricted distribution systems or, indeed, directly by the suppliers themselves. Hence, the suppliers, if they can use franchise fees to capture their vertical structure's profits, will find it in their (individual and joint) interest to employ exclusive distribution[38]. Moreover, the attractiveness of such a strategy increases with the degree of market power that the suppliers and the distributors have on their respective levels of operation.

Also exclusive dealing arrangements can be used to 'soften' competition between upstream suppliers. An argument traditionally brought forward in this context is that exclusive dealing arrangements have an effect on consumer search costs. When all suppliers use exclusive dealers and dealers are spatially differentiated, consumer search costs are higher than they would be if distributors carried the products of several suppliers instead of only one. This tends to discourage consumers from comparison shopping, limiting the extent of interbrand competition and raising industry profits. Alternatively, exclusive dealing arrangements may enhance the degree to which the

[38]The fact that the contract between a supplier and its distributor is observable to the other supplier is important, but not vital. Also when contracts are unobservable, exclusive distribution arrangements can still have an effect on the market outcome. For this to be the case, it must be that each supplier-distributor combination can somehow anticipate that the other combination uses a wholesale price that exceeds marginal cost. This will, for example, be the case when, for legal reasons or risk sharing considerations, the suppliers cannot use fixed fees to obtain profits. Similar observations can be made in the context of slotting allowances paid to retailers (Shaffer, 1991).

suppliers' products are differentiated, by combining the inherent product differentiation with the differentiation that exists at the level of the distributors. Besanko and Perry (1994) consider a model in which there are two suppliers of differentiated products and retailers are spatially differentiated. Assuming free but costly entry into retailing, they show that exclusive dealing allows the suppliers to achieve higher wholesale price margins due to, first, the absence of in-store interbrand competition and, second, the cost savings stemming from the reduction in the number of retailers of each brand that can be supported in the market. It must be said, however, that this result is somewhat biased by the underlying assumption that suppliers can only use linear wholesale tariffs. After all, this restriction leads to exactly the type of contracting inefficiency discussed by Bernheim and Whinston (1998). When suppliers can use general wholesale price schedules, exclusive dealing is likely to result less often, precisely because common representation is by itself a means to coordinate downstream competition.

In a related article, Besanko and Perry (1993) do recognize the fact that exclusive dealing may intensify competition rather than soften it. Indeed, the focus of their article is the efficiency reason for exclusive dealing identified by Marvel (1982), namely that it eliminates an upstream externality problem that may exist when suppliers provide services or make investments for distributors that they have in common[39]. The elimination of the upstream externality may lead suppliers to increase their investments in their distributors, thus intensifying competition in this dimension. By contrast, the upstream externality reduces each supplier's incentive to invest in its distributors, thus 'softening' competition. This strategic effect on the competitive equilibrium may reduce the attractiveness of exclusive dealing altogether. As a result, suppliers may wish to maintain the externality by opting for common representation[40].

The role of price restraints in softening competition has also received recent attention in the literature. Rey and Vergé (1998) study how resale price maintenance contracts that are purely vertical (i.e. not involving any explicit horizontal co-ordination) may limit both intrabrand and interbrand competition. In essence, they consider a model with two suppliers who use two common distributors (intrinsic double common agency). The existence of competition at the upstream and downstream level implies that the retail price will always be below the monopoly price and that the

[39] As discussed in section 2.2.3.

[40] Besanko and Perry's (1993) more specific results are again depending on the peculiar assumption that only linear wholesale prices can be used.

monopoly industry profit cannot be realised. Resale price maintenance may, however, be an instrument to obtain monopoly profits. Since the suppliers can catch the whole mark-up through a franchise fee, their own marginal wholesale prices have no direct effect on that profit but do influence the pricing behaviour of the competitor. Under natural conditions on the elasticities, any equilibrium (imposed) retail price is a decreasing function of the equilibrium marginal wholesale price[41]. It follows that there exists a continuum of symmetric equilibria, one for each choice of wholesale prices. In particular, there exists an equilibrium leading to the monopoly retail price[42]. The extent to which the results continue to hold when double-common agency is not an equilibrium configuration, is not entirely clear, however.

2.4.2 Facilitating cartel enforcement

Beyond 'softening' competition in the sense described in the above section, vertical restraints may also serve to facilitate outright cartel agreements. Typically, vertical restraints eliminate (or, at least, reduce), downstream competition: resale price maintenance, for example, has the effect of eliminating all price competition between distributors. As a result, competing distributors may be tempted to exert pressure on their supplier(s) to impose such vertical restraints. In this way, distributors may use vertical restraints to maintain and enforce a cartel while circumventing antitrust laws prohibiting price fixing behaviour. For such practices to occur, it does appear that the distributors must be in a particularly strong position vis-a-vis the supplier(s); it is unlikely that suppliers will generally be enthusiastic about unnecessarily high margins on the distribution level.

The use of vertical restraints, notably resale price maintenance, has also been linked to the desire of suppliers to maintain collusion, at the upstream level. Typically, individual firms have incentives to cheat on any collusive agreement so that firms must be able to monitor one another in order to collude successfully. Resale price

[41] A marginal wholesale price increase by both suppliers decreases the retail margins for both products. Since a supplier can capture the profit made by a distributor on his rival's product by means of an appropriate franchise fee, this reduces the amount of money that the supplier can raise in this way. As a result, the supplier will have a stronger incentive to lower the (imposed) retail price on its own product.

[42] The multiplicity of possible equilibria is due to the fact that suppliers have more control variables than needed. A solution to avoid this multiplicity is to introduce an effort variable which is chosen by the distributors and which affects final demand. Rey and Vergé (1998) show that the resulting unique equilibrium candidate leads to monopoly prices and profits. Further extensions involve situations in which distributors have bargaining power.

maintenance may be instrumental in this respect. Absent resale price maintenance, it may be difficult to tell if the retail price has dropped because one supplier has cheated on the collusive agreement or because one of the distributors has chosen to lower its final price on its own initiative, for example in response to local market conditions. Under resale price maintenance, final selling prices are centrally set by the respective suppliers, in which case it is simpler for these firms to monitor whether the terms of the cartel agreement are being adhered to.

A formalization has recently been provided by Jullien and Rey (2000)[43]. Their paper emphasises that, while resale price maintenance can make it easier to detect deviations from a cartel agreement, the reasons that make it hard to enforce the cartel agreement in the first place, e.g. uncertainty about demand or cost conditions, also tend to reduce the profitability of resale price maintenance. After all, if resale price maintenance is imposed on distributors, these distributors cannot use localized information, which leads to an imperfect realisation of monopoly profits[44]. In this sense, there is a trade-off between the enforceability and the profitability of the cartel agreement[45].

It has also been argued that resale price maintenance might be used to reduce the suppliers' incentives to make secret wholesale price cuts (cf. Telser, 1960). After all, when the distributors are unable to openly reduce the price at which they can sell the product to final consumers, the only way in which a lower wholesale price can induce additional sales is by indirect means (e.g. occasional secret price cuts, extra service provision, promotional activity). It must be noted that this argument is only valid when the monitoring of the collusive arrangement is watertight; if not, the supplier and the distributor could simply team up in making secret price cuts (wholesale and retail).

[43] In order to study how vertical restraints can be used to maintain collusion, Jullien and Rey (2000) build on Green and Porter (1984) and Abreu-Pearce-Stachetti (1986): the suppliers use punishment strategies in order to maintain collusion. In response to information transmitted through observed market outcomes, the punishment can be triggered or not. In the absence of resale price maintenance, retail prices react to the distributor's private information, which makes it hard to detect a deviation from collusive behaviour (a collusion game with noisy observations).

[44] The imperfect realisation of monopoly profits, in turn, increases the incentive to deviate.

[45] In addition to an effect on the detection possibilities, Jullien and Rey (2000) show that resale price maintenance also has an effect on the punishment possibilities.

2.5 Concluding remarks

In this chapter, we have elaborated upon two possible motives underlying the use of vertical agreements, the efficiency motives and the anticompetitive motives. The fact that vertical agreements are agreements concluded between companies in a vertical relationship - that is, between companies which both fulfil an indispensable function in putting the product on the market - suggests that they can often be regarded as positive: in a vertical relationship it so works out that one party will be damaged when the other party does not function properly. And 'properly' means, in by far most cases: from the point of view of the consumers, because in the end, they are supposed to buy the product. Through this special interdependent relationship, every party in a vertical agreement can, in principle, be considered a natural ally of the consumer.

However, vertical agreements can have negative consequences from the welfare point of view. When they lead to the foreclosure of markets or when they 'soften' competition, it is appropriate to act against them. The number of cases in which this will be the case is however limited: it transpires that vertical restraints are unlikely to have detrimental effects when there is no market power on either level of the industry. This conclusion works in two ways. First, when there is sufficient interbrand competition, consumers are likely to benefit from the agreement in question, as the competition which is present in the market will cause part of the possible efficiency gains to be passed on to them. As a bottom line, it is highly improbable that consumers will suffer in this case, since there are plenty of alternatives (competing products) available to them. Secondly, the analysis also shows that the strategic effects, e.g. the use of vertical agreements to soften competition, are less strong when competition in the product market is fiercer. Finally, market foreclosure is difficult to achieve when there are many players at both levels of the industry.

As for possible distinctions between the types of vertical restraints, it emerges clearly from the literature that the type of the restraint does not itself determine whether it will increase or decrease economic efficiency. As indicated by the many examples given in the above sections, a particular contract provision may have either beneficial or detrimental effects, depending on the context. This observation has led some commentators to the conclusion that the *per se* prohibition of resale price maintenance that is present in many law systems in the world is, in fact, inappropriate[46].

[46] Cf. Seabright (1998, 1999) or Van Damme (1997). Ippolito and Overstreet (1996) come to the same conclusion on the basis of an impact assessment of an FTC decision that effectively put an end

I would agree to that. Still, in my perception, there is something to say for a policy which is 'tougher' on price restraints than on non-price restraints.

The most important element in this respect is that resale price maintenance appears to provide better scope for monitoring cartel agreements than the use of exclusive territories or other non-price restraints. Resale price maintenance eliminates price variability whereas an exclusive distributor will respond to local demand and cost variations. Furthermore, from a practical point of view, it may be more feasible and cheaper to support a cartel agreement using resale price maintenance than on the basis of a rigorous and costly rearrangement of the whole distribution system[47]. It is difficult to see, for instance, music companies maintaining a cartel in compact discs on the basis of exclusive distribution systems in order to obtain the required price transparency. Finally, in terms of 'potential damage control', one could say that where resale price maintenance has a direct effect on preventing intrabrand price competition, exclusive territories work indirectly and need not mean that competition is totally ruled out, as consumers may be able to travel to other sales areas in order to purchase the product.

All in all, while an *a priori* distinction between price restrictions and non-price restrictions is theoretically not defendable and, hence, it would make more sense to 'leave all options open', resale price maintenance does seem to bear a number of risks which, in my perception, seem more difficult to check than those associated with non-price restraints. In this sense, an *a priori* distinction may be the type of 'policy guidance' (cf. section 2.1) that is required to make an antitrust policy effective in fighting collusion. As indicated, this does not mean that one should stick to a policy prohibiting resale price maintenance. In particular, one should consider lifting the ban for small market participants. Still, a policy which is generally more cautious with price restraints than with non-price restraints may be appropriate.

A final issue is the question to what extent full vertical integration (through ownership) should be looked upon differently from partial integration in the form of vertical

to the policy of price floors used by Corning Glass Works, a U.S. manufacturer of kitchenware. They conclude that in the case of Corning, an efficiency reason (the outlets hypothesis) seems a more likely explanation of its past policy of resale price maintenance than any of the anti-competitive motives.

[47] This argument is not totally adequate as the 'rigourous and costly rearrangement' may be exactly the sort of credible commitment required to sustain collusion. What is meant here are situations in which the rearrangement of the distribution system is not an option (cf. products that are typically sold in many outlets in a given city or area). It is difficult to see, for instance, music companies supporting a cartel in CDs based on exclusive distributors in order to obtain the required price transparency.

restraints. Here again, while there are many commonalities in both their intent and effect, there is one noticeable difference that merits close attention, the commitment effect. For example, the 'softening of competition' effect identified by Rey and Stiglitz (1988, 1995) that arises when suppliers (with market power) delegate their pricing decisions to distributors (with market power) depends on the supply level and the distribution level not being integrated. Similarly, suppliers may want to give their distributors more pricing flexibility in order to make them respond more aggressively against potential competitors[48]. In both examples, the vertical restraints lead to outcomes that are worse for welfare than what would be the case with full integration. In this sense, vertical restraints should not be thought of, in social welfare terms, as an intermediate between no linkage and full integration. A policy which looks with greater scrutiny to vertical restraints than to full vertical mergers need, therefore, not be inappropriate.

[48]Recall that distributors, in making their pricing decisions, do not take into account the possible effects on the profits of the upstream supplier: this is the double marginalisation problem. This may make them more aggressive when being faced by potential entry (Rey and Stiglitz, 1988, 1995).

2.6 Glossary of terms

Vertical restraints observed in practice include:

- exclusive distribution agreements: the supplier appoints a distributor for a specific territory (or, more generally, for a customer group). The supplier is typically restricted from supplying other distributors and from selling directly to customers in that area, while the distributor is generally restricted from selling into other areas[49].
- exclusive purchasing contracts/ requirements contracts: the distributor undertakes to obtain all his requirements of a specified product from the supplier.
- exclusive dealing contracts / non-compete contracts: the distributor undertakes not to sell products that compete with those of the supplier.
- selective distribution: the supplier operates a restricted system of distribution on the basis of 'approved' distributors who agree not to supply unauthorised dealers outside the network.
- quantity forcing: the distributor is required to sell at least a specified quantity of the products concerned.
- franchise agreements: the supplier (the franchisor) licenses the distributor (the franchisee) to sell goods under the supplier's trade mark, usually within a standardised and detailed commercial framework. A franchise agreement often involves the transfer of a substantial amount of know-how.
- resale price maintenance: the supplier fixes the price at which the distributor is to sell the product to the consumers.
- tie-in sales: the supplier conditions the sale to the distributor also buying other products from him.

[49]Under EU competition law, some further distinctions are made, involving, in particular, the concepts of 'active sales' into other areas and 'passive sales' into other areas; see Chapter 3.

2.7 References

Abreu, D., Pearce, D. and Stachetti, E. (1986) 'Optimal cartel equilibria with imperfect monitoring' Journal of Economic Theory 39: 251-269.

Aghion, P. and Bolton, P. (1987), 'Contracts as a barrier to entry', American Economic Review, Vol. 77: 388-401.

Baxter, W. (1990) 'Substitutes, complements and the nature of the firm' in: Trebilcock and Mathewson (eds.): The law and economics of competition policy.

Bernheim, B. and Whinston, M. (1985) 'Common marketing agency as a device for facilitating collusion' Rand Journal of Economics, 16(2): 269-281.

Bernheim, B. and Whinston, M. (1986) 'Common agency' Econometrica, 54: 923-942.

Bernheim, B. and Whinston, M. (1998) 'Exclusive dealing' Journal of Political Economy, 106(1): 64-103.

Besanko, D. and Perry, M. (1993) 'Equilibrium incentives for exclusive dealing in a differentiated products oligopoly' Rand Journal of Economics, 24(4): 646-667.

Besanko, D. and Perry, M. (1994) 'Exclusive dealing in a spatial model of retail competition' International Journal of Industrial Organization, 12: 297-329.

Besanko, D. and Perry, M. (1997) 'Resale price maintenance and service with perfectly competitive retailers' mimeo Northwestern University.

Bittlingmayer, G. (1983) 'A model of vertical restriction and equilibrium in retailing' Journal of Business 56(4): 477-498.

Bork, R. (1978) 'The antitrust paradox: a policy at war with itself', Basic Books, New York, USA.

Brennan, T. (2000) 'Vertical excuses for horizontal practices: should there be a per se antitrust rule?' Antitrust Bulletin 45(2): 467-490.

Coase, R. (1937) 'The nature of the firm' Economic 4: 386-405; reprinted in G. Stigler and K. Boulding (eds.) 'Readings in price theory' Richard D. Erwin, Homewood, Illinois, USA.

Comanor, W. (1985) 'Vertical price fixing, market restrictions and the new antitrust policy' Harvard Law Review 98: 983-1002.

Comanor, W and Frech III, H. (1985) 'The competitive effects of vertical agreements ?' American Economic Review 75(3): 539-546.

Comanor, W. and Rey, P. (1998) 'Vertical restraints and the market power of large distributors' forthcoming in Review of Industrial Organization.

Dobson, P. and Waterson, M. (1996) 'Vertical restraints and competition policy' Report for the Office of Fair Trading, December 1996.

Fasquelle, D. (1993) 'Droit américain et droit communautaire des ententes: étude de la règle de raison' Joly Editions, Paris, France.

Green, E. and Porter, R. (1984) 'Non-cooperative collusion under imperfect price information' Econometrica 52: 87-100.

Grossman, S. and Hart, O. (1986) 'The costs and benefits of ownership: a theory of vertical and lateral integration' Journal of Political Economy, 1986, vol. 94: 671-719.

Hart, O. and Tirole, J. (1990) 'Vertical integration and market foreclosure' Brookings Papers on Economic Activity: Microeconomics, p. 205-286.

Holmström, B. and Tirole, J. (1989) 'The theory of the firm' in: Schmalensee and Willig (eds.) Handbook of Industrial Organization 1, North-Holland, Amsterdam, the Netherlands.

Hovenkamp, H. (1999) 'Federal antitrust policy: the law of competition and its practice', Hornbook Series, West Publishing Co., St. Paul, Minnesota, USA.

Ippolito, P. and Overstreet, T. (1996) 'Resale price maintenance: an economic

assessment of the Federal Trade Commission's case against the Corning Glass Works' Journal of Law and Economics XXXIX: 285-328.

Jullien, B. and Rey, P. (2000) 'Resale price maintenance and collusion' mimeo Université de Toulouse.

Katz, M. (1989) 'Vertical contractual relations' in: Schmalensee and Willig (eds.) Handbook of Industrial Organization 1, North-Holland, Amsterdam, the Netherlands.

Kay, J. (1990) 'Vertical restraints in European competition policy' European Economic Review, 34(2-3): 551-561.

Krattenmaker, T. and Salop, S. (1986) 'Anticompetitive exclusion: raising rivals' costs to achieve power over price' The Yale Law Journal, 96(2): 209-293.

Martimort, D. (1996) 'Exclusive dealing, common agency, and multiprincipals incentive theory' Rand Journal of Economics 27(1): 1-31.

Marvel, H. (1982) 'Exclusive dealing' Journal of Law and Economics, 25: 1-25.

Marvel, H and McCafferty, S. (1996) 'Comparing vertical restraints' Journal of Economics and Business, 48: 473-486.

Mathewson, G. and Winter, R. (1984) 'An economic theory of vertical restraints' Rand Journal of Economics 15: 27-38.

Mathewson, G. and Winter, R. (1987) 'The competitive effects of vertical agreements: comment' American Economic Review 77(5): 1057-1062.

O'Brien, D. and Shaffer, G. (1992) 'Vertical control with bilateral contracts' Rand Journal of Economics 23(3): 299-308.

Ordover, J., Saloner, G. and Salop, S. (1990) 'Equilibrium vertical foreclosure' American Economic Review 80: 127-42.

Posner, R. (1979) 'The Chicago School of antitrust analysis' University of Pennsylvania Law Review 127: 925-948.

Reder, M. (1982) 'Chicago economics: permanence and change' Journal of Economic Literature 20: 1-38.

Reiffen, D. (1999) 'On the equivalence of resale price maintenance and quantity restrictions' International Journal of Industrial Organization 17: 277-288.

Rey, P. (1995) 'The economic analysis of vertical restraints' mimeo Université de Toulouse.

Rey, P. and Stiglitz, J. (1988) 'Vertical restraints and producers' competition' European Economic Review, 32: 561-568.

Rey, P. and Stiglitz, J. (1995) 'The role of exclusive territories in producers' competition' Rand Journal of Economics, 26(3): 431-451.

Rey, P. and Tirole, J. (1986) 'The logic of vertical restraints' American Economic Review, 76: 921-39.

Rey, P. and Tirole, J. (1996) 'A primer on foreclosure' mimeo Université de Toulouse.

Rey, P. and Vergé, T. (1998) 'Resale price maintenance and horizontal collusion' mimeo Université de Toulouse.

Phlips, L. (1995) 'Competition policy: a game-theoretic perspective' Cambridge University Press, Cambridge, United Kingdom.

Salanié, B. (1997) The Economics of Contracts: a Primer, MIT Press, Cambridge, Massachusetts, USA.

Salop, S. and Scheffman, D. (1983) 'Raising rivals' costs', American Economic Review, 73: 267-271.

Scherer, F. (1990) 'Industrial market structure and economic performance', Houghton Mifflin Company, Boston, USA.

Seabright, P. (1998) 'Competition policy towards vertical restraints: implementable

rules for competition authorities' in: Martin (ed.): Competition policies in Europe, Elsevier Science, Netherlands.

Seabright, P. (1999) 'Vertical restraints' in: Neven, Papandropoulos, Seabright: Trawling for Minnows; European competition policy and agreements between firms. Centre for Economic Policy Research, London, UK.

Segal, I. (1999) 'Contracting with externalities' Quarterly Journal of Economics 114(2): 337-338.

Shaffer, G. (1991) 'Slotting allowances and resale price maintenance: a comparison of facilitating practices' Rand Journal of Economics 22(1): 120-135.

Spengler, J. (1950) 'Vertical integration and antitrust policy' Journal of Political Economy 58: 347-352.

Telser, L. (1960) 'Why should manufacturers want fair trade?' Journal of Law and Economics 3: 86-103.

Tirole, J. (1988) The Theory of Industrial Organization, MIT Press, Cambridge, Massachusetts, USA.

Waterson, M. (1996) 'Vertical integration and vertical restraints' In: T. Jenkinson (ed.) Readings in microeconomics, Oxford University Press, UK.

White, L. (1999) 'Foreclosure with incomplete information', mimeo, Nuffield College, Oxford.

Williamson, O. (1975) 'Markets and hierarchies: analysis and antitrust implications' Free Press, New York, USA.

Williamson, O. (1985) 'The economic institutions of capitalism, Free Press, New York, USA.

Williamson, O. (1989) 'Transaction cost economics' in: Schmalensee and Willig (eds.) Handbook of Industrial Organization 1, North-Holland, Amsterdam, the Netherlands.

Chapter 3

Vertical agreements and Article 81(1) EC: on the role of economic analysis

3.1 Introduction

In the European Union a lively debate has taken place in recent years concerning the approach that should be adopted in competition policy towards vertical agreements (i.e. agreements concluded between firms operating at different levels of the production or distribution chain). Among others, this debate has centered around the question what role economic analysis should play in the application of Article 81 of the EC Treaty and, in particular, in the application of the first paragraph of the article, Article 81(1), which establishes the principle that agreements which are restrictive of competition (and which affect trade between member states) are prohibited[1].

Whereas the role of economic analysis in the application of Article 81 has by no means been a constant one throughout the years, in recent years it can be said to evolve very rapidly. On a policy level, the main driver behind this development has been the publication of the Green Paper (a consultation document) of the European Commission concerning vertical agreements in 1997, in which the subject of economic

[1]The former Article 85 was renamed to Article 81 on 1 May 1999, following the Treaty of Amsterdam. The exact wording of the Article can be found in Appendix A to this chapter.

analysis - in the sense of an analysis to ascertain whether certain business conduct leads to a reduction or increase in (consumer) welfare - took a prominent place[2] [3]. In the late 1990s, there were various reasons for the Commission to start such a policy discussion, but an important one was certainly the wish to find out whether the Commission's policy was efficient in making a distinction between the competition enhancing effects of vertical agreements and the effects restricting competition - preeminently a matter of economic analysis. For a greater part this question had been suggested by a rather general criticism on the Commission's policy, namely that it would be too formalistic and not sufficiently based on economic analysis. Many observers, from industry, national competition authorities and academic circles, pointed out that the Commission often laid too much emphasis on the legal form of contracts, on clauses, and too little on the economic impact of agreements, which an effective competition policy should be about, eventually.

Since the Green Paper, a number of steps have been taken to reform EU competition policy towards vertical restraints. On the substantive side, these steps can be seen as increasing the role for economic analysis in the application of Article 81. Before addressing the possible new role of economic analysis, it is worthwhile, however, to pay attention to the role that economic analysis has played up to now in European competition policy, for only in this way is the recent discussion understandable. And not only for this reason: it is striking to see how many points of reference particularly the jurisprudence of the European Courts[4] has offered for a greater role for economic analysis, notably under Article 81(1).

The purpose of this chapter is to provide an overview of the main developments in the interpretation of the concept 'restriction of competition' in Article 81(1) in

[2]Green Paper on Vertical Restraints in EC Competition Policy, 22 January 1997, COM(96) 721 final.

[3]Welfare can be conceived as the (weighted) sum of consumer surplus (a monetary measure of the utility derived from consumption) and producer surplus (such as profits). The weights put to consumer surplus and producer surplus imply a certain value judgment. Most attention is traditionally paid to consumer welfare. For example, the Commission itself states in its Green Paper: 'To further the interest of the consumer is at the heart of competition policy. Effective competition is the best guarantee for consumers to be able to buy good quality products at the lowest possible prices. Whenever in this Green Paper the introduction or protection of effective competition is mentioned, the protection of the consumer's interest by ensuring low prices is implied' (at paragraph 54).

[4]The European Court of Justice and, since 1988, also the Court of First Instance.

relation to vertical restraints. The focus will be on the developments in the case law of the European Court of Justice and the European Court of First Instance, as they are the most important sources of guidance. For the purpose of this analysis, it is first of all necessary to address the institutional background of the enforcement of Article 81, as it appears that this background has had a marked influence on the way in which the provisions of the article have been interpreted. Indeed, it is sometimes said that in few fields of law the manner in which the law provisions are interpreted is so much related to the question of who interprets them[5]. Secondly, it is necessary to have some understanding of the role which is attributed to competition in the EC Treaty in general. It emerges from more than 30 years of jurisprudence that the words 'restriction of competition' have been interpreted in the light of the overall objectives of the EC Treaty (in particular, the creation of a single European market), rather than in the light of competition as such. The divergence has been most apparent in the treatment of vertical restraints offering territorial protection to distributors, restraints which have - from a pure economic point of view - all too often been treated as incompatible with Article 81(1).

The chapter is organised as follows. The next section, Section 3.2, will describe the institutional background of the enforcement of Article 81 EC, whereas Section 3.3 will look into the general role attributed to competition in the EU. Section 3.4 will provide a detailed presentation of the case law of the Courts and relates this to the approach taken by the Commission. A description and evaluation of the respective policy steps since the appearance of the Green Paper will conclude this chapter.

3.2 An institutional background

The main provision in EC competition law dealing with vertical agreements is Article 81 of the EC Treaty (former Article 85). The article consists of three paragraphs[6]. The first, Article 81(1), establishes the general principle: prohibited are 'all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their

[5] Cf. Heinemann (2000).

[6] The exact wording of Article 81 can be found in Appendix A to this chapter.

object or effect the prevention, restriction or distortion of competition within the common market'. The second paragraph provides for a sanction: according to Article 81(2), any agreement or decision prohibited pursuant to Article 81 shall be automatically void, i.e. unenforceable in the courts. The third paragraph, 81(3), establishes an exemption possibility to the general prohibition principle: it states that Art. 81(1) may be declared inapplicable when the agreement (or decision or concerted practice) satisfies the criteria mentioned in paragraph 3, such as the requirement that it contributes to economic progress[7].

Central to the application of Article 81 EC is the notion of what constitutes a 'prevention, restriction or distortion of competition' under Article 81(1). Whereas Article 81(1) states a few broad examples of what might constitute such restrictions[8] (fixing prices or other terms of trade; limiting or controlling production or technical development; sharing markets or sources of supply; discriminatory practices placing firms at a competitive disadvantage; tying arrangements), these have left ample opportunity for different interpretations in the application of this article in individual cases, notably in the context of vertical agreements.

More than anything else perhaps, it turns out that the particular allocation of powers in the application of Article 81/85 has had a profound impact on the way in which vertical agreements have been assessed under the article in the past thirty years[9]. This allocation of competences was not set out in the Treaty itself, but in a subsequent Regulation adopted by the Council of Ministers, Regulation 17 of 1962 (often briefly referred to as 'Regulation 17'). This Regulation established that both the European Commission and national courts have the competence to apply Article 85(1), but that only the European Commission is empowered to grant exemptions

[7] In order to be exemptable, Article 81(3) mentions four criteria which must all be satisfied. Two of these criteria are positivily stated: the agreement must (i) contribute to improving the production or distribution process or to technical or economic progress and (ii) allow consumers a fair share of the benefit. The remaining two are negatively stated: the agreement should (iii) not involve restrictions that are not indispensable to the goal to be achieved and (iv) not allow the companies to eliminate the competition in respect of a substantial part of the products in question.

[8] See appendix. Whereas, at first sight, there would appear to be a certain difference between the terms 'prevention', 'restriction' and 'distortion' of competition, in practice little distinction has been made (cf. Ritter, 1993, p.74).

[9] For the purpose of clarity the old name 'Article 85' will be used when Article 81 is discussed in its historical context.

under Article 85(3)[10] - a division of competences which is still in force today[11]. At the origin of concentrating the competence to apply Article 85(3) in the hands of the European Commission was the idea that it was necessary for a uniform, Community-wide working of the competition rules in a period in which few member states had experience with competition law enforcement.

In addition to the above attribution of competences to apply Art. 85, Regulation 17 also set out the specific rules governing the procedures to be followed. One procedural aspect in particular created a substantial practical difference in the treatment of agreements which were legal because they did not infringe Article 85(1) in the first place and agreements found to be legal because they were exemptable under Article 85(3). Regulation 17 established an official notification system for agreements for which firms sought exemption: without such notification it was impossible to obtain exemption and the agreement, if found to infringe Article 85(1), would be automatically void[12]. Therefore, unlike agreements which did not infringe Article 85(1) in the first place, agreements caught by Article 85(1), but exemptable under Article 85(3), had to pass through the EC procedural system in order to be of any legal use for the companies involved.

Not long after the system was put in place (and the Commission was facing the prospect of having to deal with more than 30.000 notifications), the Commission took steps to streamline the procedure. Whereas Regulation 17 empowered the Commis-

[10]The EC Treaty is more than a treaty establishing mutual obligations between signatory nations (cf. Ritter e.a., 1993). In contrast to most treaties, the EC Treaty establishes a body of law that has become part of the national law of each member state and is directly enforceable by the national courts. The so-called directly applicable provisions of the Treaty create rights and obligations horizontally, among individuals, or vertically, between individuals and governments. A condition for the direct application of Community law is that the relevant provisions are 'self-executing'; they must be suited, according to their legal nature, system and wording to have direct application in the legal relationships of the Community and its subjects. In the case of Article 81, it was Regulation 17 that has made the prohibition contained in Article 85(1) directly applicable by setting out the procedures to be followed and the assignment of competences. National courts may apply Article 85(3) only to determine that the conditions for the grant of an exemption are clearly not fulfilled (Case C-234/89 *Delimitis S. v Henninger Bräu*, judgment of 28 February 1991, (1991) ECR 935). In case of doubt, both in the application of Articles 85(1) and 85(3), they are advised or, in many cases, bound to stay proceedings and to request a preliminary ruling from the European Court of Justice (a ruling on questions of law concerning the interpretation of the EC Treaty) or to await the opinion of the European Commission.

[11]Regulation 17 has not been subject to major changes, with one possible exception (a procedural change with respect to exemptions in 1999). A far reaching reform is, however, being proposed (cf. Section 3.5).

[12]For agreements not infringing Art. 85(1) an official decision to this end (a 'negative clearance') could be sought but this was not necessary.

sion to grant exemptions in individual cases, a later Council Regulation, Regulation 19 of 1965, allowed the Commission to adopt regulations (the 'block exemption' regulations) which define certain categories of agreements which generally fulfil the conditions of exemption under Article 85(3). Such agreements would be automatically exempted. The block exemption regulations in the field of distribution which were adopted pursuant to Regulation 19/65 related to exclusive distribution, exclusive purchasing, franchising and motor vehicle distribution agreements. Just as Regulation 17 provided that the European Commission could grant exemptions accompanied by conditions and obligations on the companies party to the agreement, the block exemption regulations contained provisions which stipulated what could and what could not be exempted.

The legal monopoly to apply Article 85(3), the notification system and the block exemption system gave the Commission considerable scope to adopt a regulatory approach towards Article 85. It is fair to say that the Commission has effectively taken a regulatory approach, in two respects. First, it has fairly consistently chosen for a broad interpretation of the concept 'restriction of competition' under Article 85(1). In essence, the Commission has tended to equate a 'restriction of competition' with a restriction on the economic freedom of the companies concluding the agreement or third parties[13]. As vertical agreements usually amount to some sort of restriction (e.g. a promise to grant exclusivity), many of them were prone to be caught by Article 85(1). By taking this approach, the Commission ensured that most vertical agreements ended up in the regulatory framework of the exemption system. Marenco (1999) explains this interpretation of 'restriction of competition' by noting that in judging whether or not an agreement between companies will reduce welfare ('restrict the quality or quantity of supply'), one first needs to establish whether the agreement has the tendency to do so. According to him, the 'restriction of economic freedom' lends itself pre-eminently to identify the existence of such a tendency[14]. Goyder (1993) points out in this respect that the words 'restriction of competition' have been interpreted in a way to assert EU jurisdiction over the use of vertical agreements in the common market, rather than to give an assessment of whether individual agreements can be justified. While the European Court of Justice has generally endorsed the

[13] Cf. Deacon (1995) or Hawk (1995).

[14] Marenco (1999) further relates the interpretation of 'restriction of competition' and its emphasis on the freedom of economic action to the legal tradition, i.c. the early interpretation of the concept 'restraint of trade' in the anglo-saxon common law systems. See also Appendix B to this chapter.

Commission's application of Article 85(1), it has in some cases placed limits upon it, as will be discussed in detail in Section 3.3.

Perhaps the clearest illustration of the Commission's approach can be given in the context of the early cases involving exclusive distribution agreements. Rather than looking into actual anti-competitive effects or issues of market power, the Commission would typically find that the appointment of an exclusive distributor for a sales area constituted a 'restriction of competition' on the basis that the supplier thereby deprived himself of the ability to appoint other distributors in the area (for the duration of the contract) and that other distributors were prevented from sourcing the products in question directly from the supplier[15]. At the same time, the Commission normally granted exclusive agreements an exemption under Article 85(3) as it held the opinion that they are 'often the most effective way and sometimes indeed the only way for the manufacturer to enter the market and compete with other manufacturers already present'[16]. But if a distribution system involving exclusive distribution is the most effective way or the only way for a manufacturer to enter the market and to compete with other manufacturers already present, how can it be that the same distribution system amounts to a 'restriction of competition' under Article 85(1) in the first place?

The second respect in which the Commission adopted a regulatory approach was that the block exemption regulations were drafted in a rather detailed way, with provisions describing which clauses should not be in an agreement for it to be exempted (the 'black clauses') and which clauses could be in (the 'white clauses'). These regulations offered little flexibility and were considered to have a straight-jacket effect, in particular in view of the fact that agreements governing comparable economic relationships as those covered by the block exemptions but which did not precisely match the terms of the regulations did typically not benefit from the block exemptions[17]. Barendrecht and Van Peursem (1997) mention the example of an exclusive

[15]Cf. Commision Decisions *Grundig/Consten* (Official Journal 1964, 2545), *DRU/Blondel* (OJ 1964, 2194), *Hummel/Isbecque* (OJ 1965, 2581), *Jallatte/Voss* (OJ 1966, 37) or, in the context of exclusive license agreements, *Nungesser* (OJ 1978, 1286). Similar remarks can be made with regards to manufacturers using different wholesale prices, depending on the final destination of the merchandise (*Distillers*, Official Journal 1978, L50/16).

[16]Regulation 1983/83 on the application of Article 85(3) to categories of exclusive distribution agreements, recital 7 (Official Journal 1983 L 173/1). By contrast, as will be explained below, agreements preventing parallel imports into the sales area (i.e. imports not going through the 'official' channels) have never been exempted.

[17]This could even be the case when the clauses which caused the agreement to fall outside the relevant block exemption regulation did not infringe Article 85(1) in the first place (Goyder, 1996).

distribution agreement not enjoying the benefit of the corresponding block exemption regulation (and therefore being considered void under Article 85(2)) merely because the agreement contained a clause giving the manufacturer the right to buy back all remaining stocks after termination of the agreement[18]. More importantly perhaps, the fact that there existed block exemption regulations for certain types of distribution agreements (e.g. franchise systems[19]) but not for others (e.g. selective distribution systems[20]) meant that companies tended to frame their distribution systems into those specifically approved by one of the block exemption regulations.

Against the background of these institutional arrangements for the enforcement of Article 85 as a whole, the role of economic analysis in the application of Article 85(1) - in the sense of Chapter 2: analysing the effects enhancing competition and those restricting competition - has been fairly limited. Rather, the main accent of economic analysis within competition policy lied in the application of Article 85(3), the exemption procedure. This does not mean, however, that the broad interpretation of the concept 'restriction of competition' under Article 85(1) was, in some way, an arbitrary choice made without reference to any competition theory. The Commission's thinking refers, of course, to the classic notion of competition as a process of rivalry among economic agents who act independently[21]. It also refers, however, to a specific school of thought which has been particularly influential in the development of post-war competition policy in Europe, the Freiburg School[22]. This school of thought will be briefly discussed in the next section.

[18]Barendrecht and Van Peursem (1997) discuss this example in their comment on the judgment of the Court of Rotterdam in the case *Novem/Johnson*, 11 March 1994, rolnr. 5748/92.

[19]In a franchise agreement, the supplier (the franchisor) licenses the distributor (the franchisee) to sell goods under the supplier's trade mark, usually within a standardised and detailed commercial framework. A franchise agreement often involves the transfer of a substantial amount of know-how.

[20]Selective distribution systems are distribution systems on the basis of 'approved' distributors who agree not to supply unauthorised dealers outside the network.

[21]Cf. Van den Bergh (1997). The classic notion of competition relates to the behaviour of individual sellers aiming to sell their merchandise by offering better conditions than their rivals (or buyers offering better prices than rival buyers). Competition was viewed as a force which would lead individually acting persons to take actions in line with the common interest (cf. Adam Smith's (1776) 'invisible hand').

[22]See Deacon (1995), Van Gerven e.a. (1997) or Fasquelle (1993).

3.3 On the role of competition

Conceptually, one can look at the role of competition in free market economies in two different ways. One may regard competition as being an (almost) absolute condition for economic prosperity and competition as being a means to achieve it (Fasquelle, 1993)[23].

In the first conception, competition is viewed as the essential carrier of economic progress and, indeed, a prerequisite for economic progress. Accordingly, the legal system should aim at repressing all market arrangements and forms of market behaviour which lead to a reduction of competition. In such a case, competition policy is all about determining the criteria that allow for the identification of market arrangements and behaviour which are in restraint of competition.

In the second conception, competition is equally looked upon as the ordinary way to achieve economic progress, but it is not an absolute priority. As such, competition is a means to foster economic progress, but possibly one among many. By consequence, a more general value judgement is appropriate for market arrangements and market behaviour.

The notion of a social market economy (*soziale Marktwirtschaft*) elaborated by German scholars closely related to the University of Freiburg probably provides one of the clearest examples of the second way of perceiving competition, i.e. as a means rather than a condition. Leading representatives of the so-called Freiburg School[24], such as Eucken, Böhm and Müller-Armack, emphasized that the specific way in which an economic process develops is dependent upon the kind of economic system that prevails (the property right system, the monetary system, the organisation of markets, ...)[25]. In the view of the Freiburg School, the free market economy and the competitive process should be understood as elements of an economic constitution, the object of which is to achieve sustained economic performance and stability. According to

[23]This section draws upon Fasquelle (1993), Souty (1996) and Hildebrand (1998).

[24]The Freiburg School is also known as the 'ordoliberal' school of thought.

[25]Many leading German political figures were associated with the Freiburg School. Ludwig Erhard, a student of Walter Eucken, was one of the architects of German post-war economic policy, notably in his function of Minister of Economic Affairs and, later, Chancellor. During his term, he was one of the driving forces behind the 1957 Act against Restraints of Competition (Gesetz gegen Wettbewerbsbeschränkungen, GWB). Similarly, Walter Hallstein (one of the founders of the European Communities and the first president of the European Commission) and Hans von der Groeben (one of the drafters of the 'Spaak Report' and the first Commissioner for competition policy of the European Commission) are often associated with the Freiburg School.

Eucken: 'This principle of the economic constitution does not only require that certain measures of economic policy are avoided, such as public subsidies, the creation of state monopolies, the control of prices or import prohibitions. Neither is it by itself sufficient to prohibit cartels. The principle is not in the first place a negative one. What is necessary is an economic policy that aims at the development of competition in all domains and that hence contributes to the fundamental principle of an economic constitution'[26]. The Freiburg School agrees with earlier conceptions of liberalism in considering a competitive economic system to be essential for a prosperous, free and equitable society. Nevertheless, it added new legal and social dimensions to the liberal tradition. As such, the Freiburg School clearly viewed competition as an important element but placed it in a wider perspective.

As for the European context, it is clear that the notion of 'competition as an instrument' is the one reflected in the EC Treaty (Hawk, 1989; Fasquelle, 1993). Articles 2 and 3 EC set out the principles of the Treaty. Under Article 2, the Community has as its task, by establishing a common market and progressively approximating the economic policies of member states, to promote throughout the Community a harmonious development of economic activities, a continued and balanced extension, an increase in stability, an accelerated raising of the standard of living and closer relations between the member states. One of the means of obtaining these goals, set forth in Article 3(g) of the Treaty, is 'the institution of a system ensuring that competition in the common market is not distorted'[27]. Other instruments mentioned in the Article include the co-ordination of the economic policies of the member states, the establishment of a common external trade policy and a common agricultural policy.

In the European context, therefore, it appears that fostering competition is not an end in itself, but a means to achieve the broader objectives of the EC Treaty. Several opinions can be found in the literature as to what the main objectives and instruments exactly are. However, when it comes to comparing competition and market integration (in the sense of the free movement of goods, services, capital and persons), the latter clearly stands out as the most important instrument. Deacon (1995) and Paulis and Peeperkorn (1998) in this respect even speak about a 'quasi droit acquis' or 'inalienable right' of consumers or their agents to purchase wherever

[26] W. Eucken 'Grundsätze der Wirtschaftspolitik' Bern-Tübingen, 1952. Reproduced in Souty (1996).

[27] Article 3(g) was renamed to Article 3(1)(g) on 1 May 1999, following the Treaty of Amsterdam. In the original EC Treaty it was Article 3(f).

they want. This, of course, refers to a particular interpretation of the concept of market integration, but the principle is clear[28].

As to competition policy itself, the Freiburg School had a fairly tough attitude in the sense that the model of perfect competition (diffusion of market power) was regarded as the appropriate substantive standard for competition law (Van den Bergh, 1997). According to their representatives, history had demonstrated that competition tended to collapse because enterprises preferred private (i.e. contractual) regulation of business activities rather than competition and because enterprises were frequently able to acquire such high levels of economic power that they could eliminate competition (Hildebrand, 1998). Competition law was viewed as a means of preventing this degeneration of the competitive process. In line with its reservations to private (contractual) regulation of business activities as opposed to (perfect) competition, the Freiburg School has tended to equate restrictions of economic freedom with 'restrictions of competition'. This strict interpretation of the concept 'restriction of competition' is also the one that has been adopted in European competition policy (Deacon, 1995; Hawk, 1995), though more in the approach of the European Commission than in that of the European Courts - as will show from the next sections[29].

3.4 The European Courts on Article 85(1)

As indicated, Article 81(1) EC - formerly Article 85(1) - prohibits all agreements and concerted practices 'which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the common market'. In this section a number of leading cases will be discussed in order to obtain an understanding of what the European Courts have considered to amount to a 'restriction of competition'. Unfortunately, as Goyder (1993) points out, it is difficult to refer to a 'rule' in the interpretation of the term 'restriction of competition' in the context of Article 85(1). Rather, '[w]hat is important for a clear understanding of these words is knowledge of the way in which the Commission and, more importantly, the European Court and the Court of First Instance have interpreted them in a variety

[28]The next section, Section 3.4, will further explore the relation between the two instruments (competition and market integration) in EC policy, in the context of a number of specific competition cases.

[29]It is worth noting that the strict interpretation of the concept 'restriction of competition' has been only partially retained in the German Act against Restraints of Competition (GWB), as is illustrated by the distinct treatment of vertical and horizontal agreements under this law.

of different contexts'[30]. It emerges from the review of cases that a number of phases and approaches can be identified, starting from the early cases of the 1960s to the introduction of the notion of 'appreciability', the introduction of the European 'rule of reason' and the development of the so-called cumulative effect doctrine. In the below sections, the various approaches will be presented and compared to one another, as well as to the economic test of balancing the positive and negative consequences of vertical restraints on competition.

3.4.1 The early cases of the 1960s

A first case relevant to the interpretation of the expression 'restriction of competition' in Article 85(1) is *Société Technique Minière (STM) v. Maschinenbau Ulm* of 1966[31]. This was a case in which the Court of Justice was asked by a French court to give a preliminary ruling (a ruling on questions of law concerning the interpretation of the EC Treaty) on a case between the above mentioned parties[32]. The case concerned an exclusive supply contract, by which STM enjoyed the exclusive rights to sell certain equipment produced by Maschinenbau Ulm in France. The contract did not insulate the French territory, STM could sell the goods outside France and parallel imports could be obtained from other countries. In its ruling, the Court started off by indicating the general principle that

> 'In order to be prohibited as being incompatible with the common market under Article 85(1) of the Treaty, an agreement between undertakings must fulfil certain conditions depending less on the legal nature of the agreement than on its effects on 'trade between member states' and its effects on 'competition'. Thus as Article 85(1) is based on an assessment of the effects of an agreement from two angles of economic evaluation,

[30] Goyder (1993).

[31] Case 56/65 *Société Technique Minière v. Maschinenbau Ulm*, judgment of 30 June 1966, (1966) ECR 235.

[32] As the name indicates, a preliminary ruling of the European Court is not a final ruling on the case itself (*in casu* the dispute between STM and Maschinenbau Ulm). Rather, it is a ruling on questions of law asked by the national court concerning the interpretation of the EC Treaty. After the preliminary ruling of the European Court, it is up to the national court to decide on the outcome of the case itself, in the light of the clarifications given by the European Court. In the literature on European competition law, much attention is paid to the preliminary rulings, but the final rulings tend to be ignored. This is a pity, since the final rulings are part of EC law just the same.

> it cannot be interpreted as introducing any kind of advance judgment with regard to a category of agreements determined by their legal nature'.

In this ruling, it became clear that the reference made in Article 85(1) to the words 'object' and 'effect' is of particular importance. First, the Court specified that for an agreement to be caught by the prohibition contained in Article 85(1) it must have either as its object or as its effect a restriction of competition. Consequently, these are alternative and not cumulative requirements. The Court also appeared to make a distinction as to the factual elements which are necessary to establish that an agreement has the object or effect of restricting competition.

> 'The fact that these are not cumulative but alternative requirements, indicated by the conjunction 'or', leads first to the need to consider the precise purpose of the agreement, in the economic context in which it is to be applied. This interference with competition referred to in Article 85(1) must result from all or some of the clauses of the agreement itself. Where, however, an analysis of the said clauses does not reveal the effect on competition to be sufficiently deleterious, the consequences of the agreement should then be considered (...)'.

The Court has clarified in later rulings that this means that if the purpose (the object) to restrain competition is evident, then the agreement itself, or at least any anti-competitive clause in the agreement, constitutes a restriction of competition 'by its very nature'[33]. In such a case, there is no need to take further account of the concrete effects of an agreement for the purposes of applying Article 85(1). The analysis of the effect, on the other hand, is designed to establish whether an agreement whose purpose is not anti-competitive, is nevertheless liable to affect competition in the specific market context. The Court further specified that:

> 'The competition in question must be understood within the actual context in which it would occur in the absence of the

[33]Joined cases 56/64 and 58/64 *Consten and Grundig v. Commission* of 13 July 1966, (1966) ECR 299. Case 19/77 *Miller International Schallplatten GmbH v Commission* of 1 February 1978, (1978) ECR 131. Case C-306/96 *Javico International and Javico AG v Yves Saint Laurent Parfums SA* (1998) ECR I-1983. It follows from these cases that the concepts of restriction by object and restriction by nature are synonymous (Woods and Filipponi, 1999).

> agreement in dispute. In particular it may be doubted whether there is an interference with competition if the said agreement seems really necessary for the penetration of a new area by an undertaking. Therefore, in order to decide whether an agreement containing a clause 'granting an exclusive right of sale' is to be considered as prohibited by reason of its object or of its effect, it is appropriate to take into account in particular the nature and quantity, limited or otherwise, of the products covered by the agreement, the position and importance of the grantor and the concessionnaire on the market for the products concerned, the isolated nature of the disputed agreement or, alternatively, its position in a series of agreements, the severity of the clauses intended to protect the exclusive dealership or, alternatively, the opportunities allowed for other commercial competitors in the same products by way of parallel re-exportation and importation'.

In principle, therefore, the Court stated that some sort of economic analysis should play a role in establishing whether an agreement has as its object or its effect a restriction of competition. It must be said however that, in practise, the Court contented itself with much more limited analyses than what might be expected from the above quotation, especially in the assessment of restrictions by object (Fasquelle, 1993)[34]. This appears clearly in another landmark case of 1966, *Consten and Grundig*[35].

In *Consten and Grundig*, the Court had to a rule on an appeal against a Commission decision concerning a distribution agreement between Grundig, the German manufacturer of radio and television sets, and Consten, a French wholesaler. Grundig had appointed Consten as its exclusive distributor in France. In turn, Consten agreed not to sell goods which competed with the Grundig range of products and also agreed that it would not make deliveries into other sales territories, after having received an assurance from Grundig that similar restrictions had been placed on Grundig dis-

[34] As indicated before, a preliminary ruling of the European Court is not a final ruling on the case itself. I am not aware of the final ruling of the French court in the case *STM v. Maschinenbau Ulm* or of any literature discussing the final outcome.

[35] Joined cases 56/64 and 58/64 *Consten and Grundig v. Commission*, judgment of 13 July 1966, (1966) ECR 299.

tributors in other countries. Following a complaint made by a French distributor[36], the Commission issued a decision in which it declared that the agreement violated Article 85(1) on the basis that the arrangements were intended to protect Consten from competition by other distributors of Grundig products. As a matter of fact, the Commission had completely focused its examination on the distribution system of Grundig products, i.e. on intrabrand competition. The Commission was of the opinion that as customers may find it difficult to compare prices for branded products such as Grundig, intrabrand competition is important for preserving the possibility for customers to obtain low cost supplies of Grundig products.

Consten and Grundig appealed against the Commission's decision at the European Court. During the proceedings, the German government intervened in support of the arguments made by Grundig and Consten. While acknowledging that the agreement restricted competition between the distributors of Grundig products, the applicants and the German government maintained that the main source of competition was to be found at the level of the producers of different brands. Specifically, they were of the opinion that 'since the Commission restricted its examination solely to Grundig products the Decision was based upon a false concept of competition and of the rules on prohibition contained in Article 85(1), since this concept applies particularly to competition between similar products of different makes; the Commission, before declaring Article 85(1) to be applicable, should, by basing itself upon the 'rule of reason', have considered the economic effects of the disputed contract upon competition between the different makes'.

The 'rule of reason', to which the applicants and the German government referred, is a concept that takes a central position in American antitrust law. In American antitrust, the rule of reason entails an assessment of the *net effect* of the agreement on competition, implying a comparison between the pro-competitive and anti-competitive effects in the light of the prevailing market circumstances, without reference to any other elements having no bearing with competition (Van Gerven e.a., 1997). As such, the American rule of reason is about 'balancing' the positive and negative effects on competition. In this balancing, interbrand competition - competition at the level of

[36]The French distributor, UNEF (a predecessor of the well-known French retail chain FNAC), was denied first-tier access to Grundig products, as a result of the exclusive relationship between Consten and Grundig.

the producers - has tended to dominate the analysis[37].

In addition, Grundig and Consten (again supported by the German government) held the view that there is a 'presumption that vertical sole distributorship agreements are not harmful to competition and in the present case there is nothing to invalidate that presumption. On the contrary, the contract in question has increased the competition between similar products of different makes'. The applicants in this respect referred to the 'free rider' argument[38] and the need to protect investments as a justification for the use of exclusive territories.

Advocate-General Roemer, who delivered his opinion on the case, was also very critical of the approach taken by Commission: 'Doubtless it is undeniable that in a given market situation competition between several sellers of a single product can also take on great importance, that it may be indispensable for the normal play of competition on the market. But the Commission is wrong in taking account of this last-mentioned internal competition exclusively and in neglecting completely in its considerations competition with similar products. In fact, it is perfectly possible that there exists between different products or rather between different producers such sharp competition that there remains no appreciable margin for what is called internal competition in a product (for example in relation to price and servicing). The Commission considers that it does not have to take into consideration this competition between different manufacturers except for simple mass product articles. That does not seem to be correct, if it is desired to judge economic phenomena realistically. It must first be said that the Commission came to its conclusion on the export prohibition after considering only its abstract aim, without examining its concrete repercussions on the market. In so doing, in my opinion, it did not apply Article 85(1) correctly'.

[37] See Appendix B to this chapter for a fuller presentation of the American rule of reason. Two remarks can be made at this point. First, the rule of reason is often put as the opposite of a *per se* rule. In the US context, this is not entirely accurate (Fasquelle, 1993; Hawk, 1995). Rather, the *per se* approach is nothing but an aspect of the rule of reason in its proper sense (i.e. balancing the effects on competition), special to certain types of agreements which are regarded to be usually detrimental to competition. Second, as regards the balancing of the effects on competition: it is only since the *Sylvania* decision of 1977 that economic efficiency became the (main) standard of competition law (in line with the ideas of the Chicago School). Before that, the stricter Harvard views prevailed.

[38] Cf. Chapter 2. When several retailers sell the same product, every retailer will benefit from it when other retailers make efforts to advertise the product or promote it by giving pre-sale advice. In such cases the incentives for retailers to make these kinds of effort may be too low (from the viewpoint of the vertical structure as a whole) as competitors which do not make these expenses can accordingly compete with sharp prices, while at the same time benefiting from the efforts made by other retailers.

The Court of Justice, however, endorsed the Commission's approach. First of all, it recalled what it had already indicated in earlier decisions (such as the above mentioned *STM v. Maschinenbau Ulm*), namely that Article 85(1) refers both to agreements between parties at the same level in the economy (horizontal agreements) and agreements between parties at different levels (vertical agreements). In particular, it said:

> 'The principle of freedom of competition concerns the various stages and manifestations of competition. Although competition between producers is generally more noticeable than that between distributors of products of the same make, it does not thereby follow that an agreement tending to restrict the latter kind of competition should escape the prohibition of Article 85(1) merely because it might increase the former'.

So, the Court was very aware that vertical agreements could work out very well for competition, in this case for competition between brands. Nevertheless it ruled that the concept of 'competition', which is not defined in further detail in Article 85, should be understood as relating to competition at the distribution level as well. As a matter of fact, the emphasis of the judgment was laid on exactly this type of competition; the very purpose of the distribution agreement between Grundig and Consten, preventing competition by other importers, led the Court to prohibit the agreement under Article 85(1). The Court did not hide one of its prime motivations for taking this tough line:

> '[A]n agreement between producer and distributor which might tend to restore the national divisions in trade between Member States might be such as to frustrate the most fundamental [objectives] of the Community. The Treaty, whose preamble and content aim at abolishing the barriers between states, and which in several provisions gives evidence of a stern attitude with regard to their reappearance, could not allow undertakings to reconstruct such barriers. Article 85(1) is designed to pursue this aim (...)'.

Because the Court thus made use of the freedom of interpretation offered by Article 85, distribution agreements tending to restore the compartmentalization of markets

along national lines came, in principle, to fall under Article 85(1) and that was also what the Court was after. The priority of Community policy was on lowering trade barriers, both those erected by member states and those erected by private parties, and the private barriers were especially found at the level of resale. Still, even from the viewpoint of market integration, there is much to say for the arguments brought forward by the parties, in particular the free-rider argument. After all, if entering a new (national) market requires substantial investments from the side of a local distributor, it may be necessary to grant this distributor protection against intrabrand competition (e.g. from parallel importers)[39]. Woods and Filipponi (1999) note, however, that one should not forget that in the 1960s progress towards the establishment of the single European market was nowhere near as advanced as it is today. For example, the landmark decisions of the Court of Justice on the free movement of goods in the *Dassonville* and *Cassis de Dijon* cases only occurred in the mid to late 1970s[40]. In a sense, the Court still had to build a 'culture' of European integration, an objective perhaps best pursued by recognizable policies such as the one chosen in *Consten and Grundig*.

As for the necessary analysis to be performed, the Court confirmed its position that for the purpose of applying Article 85(1), there is no need to take account of the concrete effects of an agreement once it appears that it has as its object a restriction of competition (in the sense of EC competition law).

> 'Therefore the absence in the contested decision of any analysis of the effects of the agreement on competition between similar products of different makes does not, of itself, constitute a defect in the decision. It thus remains to consider whether the

[39] Cf. Sections 2.2.2 (on controlling externalities between distributors) and 2.2.5 (on transaction costs).

[40] Case 8/74 *Procureur du Roi v Benoit and Gustave Dassonville*, judgment of 11 July 1974, (1974) ECR 837; case 120/78 *Rewe Zentral AG v Bundesmonopolverwaltung für Branntwein ('Cassis de Dijon')*, judgment of 20 February 1979, (1979) ECR 649. In *Dassonville*, the Court determined that all trading rules enacted by member states which are capable of hindering, directly or indirectly, actually or potentially, intra-community trade are to be considered as measures having an effect equivalent to quantitative restrictions in the sense of Article 30 EC (nowadays Article 28) and thereby, in principle, forbidden. In *Cassis de Dijon* the Court narrowed down the exception possibility provided by Article 36 EC (now Article 30), by establishing the principle of mutual recognition of national product regulations concerning, for instance, food safety. These two judgments constituted a breakthrough in a context in which national product regulations often acted as barriers to trade and EU harmonization of such regulations was a very slow process (cf. Kapteyn and VerLoren van Themaat, 1995).

> contested Decision was right in founding the prohibition of the disputed agreement under Article 85(1) on the restriction on competition created by Grundig products alone'.

In the following, the Court did mention the market context and specified on one occasion that Grundig is 'a very well-known brand'. However, rather than relating this observation to the effect that the agreement might have on interbrand competition, it used it as an argument in support of the Commission's choice to restrict its analysis to intrabrand competition (within the Grundig distribution system), basing itself on the same arguments as the Commission had used. As the Commission had clearly established that the distribution system aimed at isolating the French market for Grundig products, the Court found that it was therefore proper for the contested decision to hold that the agreement constituted an infringement of Article 85(1). The Court added:

> 'No further considerations (...) and no possible favourable effects of the agreement in other respects, can in any way lead, in the face of the abovementioned restrictions, to a different solution under Article 85(1)'.

It would appear, therefore, that the Court in this case did not apply the detailed indications that it had given in *STM v. Maschinenbau Ulm* (even though these appeared to relate to restrictions by object as well[41]), but simply categorised absolute territorial protection as a restriction of competition. This supports the hypothesis that the standard of proof is different for restrictions by object than for restrictions by effect[42]. According to Hildebrand (1998), the case *Consten and Grundig* should

[41]Recall that in *STM v. Maschinenbau Ulm* the Court stated: 'Therefore, in order to decide whether an agreement containing a clause 'granting an exclusive right of sale' is to be considered as prohibited *by reason of its object or of its effect*, it is appropriate to take into account in particular the nature and quantity, limited or otherwise, of the products covered by the agreement, the position and importance of the grantor and the concessionnaire on the market for the products concerned, the isolated nature of the disputed agreement or, alternatively, its position in a series of agreements, the severity of the clauses intended to protect the exclusive dealership or, alternatively, the opportunities allowed for other commercial competitors in the same products by way of parallel re-exportation and importation' (emphasis added).

[42]In this respect, one should also keep in mind that the circumstances of the two cases were different. Whereas the case *STM v. Maschinenbau Ulm* concerned the attribution of an exclusive distribution right as such, in *Consten and Grundig* the distributors had, in addition, committed to prevent trade between member states in the products concerned (Van Gerven e.a., 1996).

not be perceived as rejecting economic analysis within Article 85(1), but rather as indicating that such an analysis cannot serve to validate absolute territorial protection. This is consistent with the idea that, in the EC Treaty, competition is to be considered a means rather than an end and that market integration, in the sense of the free movement of goods, is given priority (see Section 3.3).

In a later decision, *Brasserie de Haecht* of 1967, the Court re-emphasized that in judging the effects of vertical agreements one should take into account the legal and economic context within which these agreements apply[43]. The Court, being asked to give a preliminary ruling on the subject of exclusive dealing agreements, specified that where agreements combine with others, they may have a cumulative effect on competition.

> 'In fact, it would be pointless to consider an agreement, Decision or practice by reason of its effects if those effects were to be taken distinct from the market in which they are seen to operate and could only be examined apart from the body of effects, whether convergent or not, surrounding their implementation. Thus in order to examine whether it is caught by Article 85(1) an agreement cannot be examined in isolation from the above context (...). The existence of similar contracts may be taken into consideration for this objective to the extent to which the general body of contracts of this type is capable of restricting the freedom of trade'

3.4.2 The notion of appreciability

The Court also applied its reference to the surrounding market context to agreements of minor importance. In the case *Völk/Vervaecke* of 1969 it established the requirement of appreciability[44]. This case was about an exclusive distribution agreement between Völk, a small German manufacturer of washing machines, and the Belgian distributor Vervaecke, by which the latter was granted absolute territorial protection in Belgium and Luxemburg. After stating that the question whether an agreement has as its object or effect the restriction of competition and whether it may affect trade

[43] Case 23/67 *Brasserie de Haecht v Wilkin*, judgment of 12 December 1967, (1967) ECR 407.

[44] Case 5/69 *Völk/Vervaecke*, judgment of 9 July 1969, (1969) ECR 295. As explained below, the first traces of this requirement were already present in *STM v. Maschinenbau Ulm*.

between member states should be understood by reference to the actual circumstances of the agreement, the Court concluded that

> 'an exclusive dealing agreement, even with absolute territorial protection, may, having regard to the weak position of the persons concerned on the market (...), escape the prohibition laid down in Article 85(1)'.

The Court itself did not elaborate on the relation between this judgment and its judgment in *Consten and Grundig*, where it had held that there is no need to take account of the concrete effects of an agreement once it appears that it has the restriction of competition as its object. Both the Commission, who in fact supported the idea of leaving minor agreements out of the scope of Article 85, and the Advocate-General referred in their submissions to the fact that the Court had referred in *STM v. Maschinenbau Ulm* to the effect on competition being 'sufficiently deleterious' (in the context of restrictions by object) or 'appreciable' (in the context of restrictions by effect). According to the Advocate-General, this amounted to saying that the change in competition must not be merely theoretical, but must be fairly widespread. In this context, he recalled that the Court had mentioned in *Consten and Grundig* that Grundig is a very well-known brand. As indicated, the Court itself did not further comment on this.

The requirement of appreciability has become known as the *de minimis* approach[45]. At first sight, one might interpret this approach as the reflection of some substantive economic reasoning, namely that agreements concluded between parties with little market power cannot be endangering the process of competition. This, however, appears not to have been the background (Fasquelle, 1993; Ritter e.a., 1993). The requirement of appreciability was not intended to be a substantive criterion for the appraisal of vertical agreements, but rather to be a jurisdictional criterion, i.e. to set

[45]The Commission would soon after, in 1970, issue a Notice concerning Decisions and Practices of Minor Importance, in which it indicated that, in principle, it would consider agreements between companies with a market share below 5% (and not too high a turnover) as not appreciably restricting competition. In the latest revision of this Notice in 1997, the turnover threshold was abandoned and the market share figure was increased towards 10% for vertical restraints. The Notice indicates, however, that it cannot be ruled out that agreements which are below the thresholds may still fall under Article 85(1). This is so in the case of 'particularly serious restrictions on competition' (such as restrictions on parallel imports and resale price maintenance). Nonetheless, plainly insignificant agreements of this type can escape the prohibition of Article 85(1) as well, in view of the *Völk/Vervaecke* judgment.

the limits for the application of Community law in the context of agreements between companies. Even though the contested agreement constituted by its character a restriction of competition, its market dimension was such that the Court said: this is not a case for the Community, but for the member states[46].

In this perspective, the requirement of appreciability does not enter into the heart of the concept 'restriction of competition' but is rather an element in the margin. This is not to say, however, that later commentators (e.g. Hawk, 1995) have not identified the concept of appreciability as a possible vehicle for the introduction of substantive economic analysis into Article 85(1). Before elaborating on this, we will first describe another major innovation in European competition law, namely the introduction of a 'rule of reason' under Article 85(1).

3.4.3 The European 'rule of reason'

In 1977, the Court issued the judgment *Metro I*, the first judgment about the validity of selective distribution systems (systems on the basis of 'approved' retailers)[47]. SABA, a German producer of hi-fi equipment and television sets, had established a distribution network in Germany and other European countries which involved the appointment of specialist dealers only. Department stores and discounters could, therefore, not obtain appointment. The necessary qualifications related to the nature of the premises, the training of the staff and the acceptance of substantial sales targets. Furthermore, those selling SABA products undertook not to supply resellers outside the SABA system.

Clearly there is an issue of restricting the freedom of economic action of the contracting parties within a selective distribution system as appointed dealers are allowed

[46]Furthermore, in subsequent judgments the appreciability requirement appears not to have been consistently applied to both the 'trade' component of Article 85 and the 'competition' component. According to Woods and Filipponi (1999), the appreciability of the restriction appears to be assumed in the case of restrictions by object, with the analysis of appreciability in such cases concentrating upon the effect on trade between member states. This thesis finds support in case 19/77 *Miller International Schallplatten GmbH v Commission* of 1978 (1978) ECR 131, in which the Court said that 'by its very nature, a clause prohibiting exports constitutes a restriction on competition' and in which it discussed appreciability only in the context of trade between member states. Not all commentators agree with this view, however (Faull and Nikpay, 1999). Ritter (1993) observes that the risk that the court will find a restriction to have appreciable effects on trade and competitive conditions rises in direct proportion to the anti-competitive nature of the restriction.

[47]Case 26/76 *Metro SB Großmärkte GmbH & Co KG v Commission*, judgment of 25 October 1977, (1977) ECR 1875.

to supply goods only to end-users or to other appointed dealers. That is why the Commission had concluded that the distribution system was caught by 85(1)[48]. Still, as it also considered that the system led to an improvement in distribution, it had granted an exemption under Article 85(3). It was against this decision that Metro, a German cash and carry wholesaler, brought an appeal to the European Court.

In its judgement, the Court started from the consideration that selective distribution systems constitute, together with others, an aspect of competition which accords with Article 85(1), provided that resellers are chosen on the basis of objective criteria of a qualitative nature (relating to the technical qualifications of the reseller and his staff and the suitability of his trading premises) and that such conditions are laid down uniformly for all potential resellers and are not applied in a discriminatory fashion. It inferred from this that objective restrictions which are *necessary* for the system to be realized and maintained, such as the above mentioned qualitative criteria and the obligation on resellers not to sell to non-approved resellers, are no restrictions of competition in the meaning of 85(1). This approach, which implied a departure from the Commission's rigid interpretation of the concept of 'restriction of competition', has become known as the European 'rule of reason'[49]. It refers to the analysis of restrictive clauses to ascertain whether they are necessary to secure the realisation or implementation of an agreement which 'accords with Article 85(1)'.

In explaining its choice, the Court referred for the first time to the notion of 'workable competition':

> 'The requirement contained in Articles 3 and 85 of the EC Treaty that competition shall not be distorted implies the existence on the market of workable competition, that is to say the degree of competition necessary to ensure the observance of the basic requirements and the attainment of the objectives of the Treaty, in particular the creation of a single market achieving conditions similar to those of a domestic market. In accordance

[48] To be precise, the Commission held that the obligations on wholesalers and dealers not to supply dealers not approved by SABA had 'as their object and effect the restriction of competition within the common market to a perceptible degree'. By contrast, to the extent that admission to the system was based on general, justifiable qualitative criteria and all dealers fulfilling these criteria were in fact appointed as SABA dealers, no restriction of competition within the meaning of Article 85(1) would arise. Commission Decision *SABA*, Official Journal 1976, L 28, p.19.

[49] As will be discussed later, the European rule of reason is markedly different from the rule of reason as it is applied in US antitrust. See also Appendix B to this chapter.

> with this requirement the nature and intensiveness of competition may vary to an extent dictated by the products or services in question and the economic structure of the relevant market sectors'.

More specifically on selective distribution systems the Court indicated that:

> 'It is true that in such systems of distribution price competition is not generally emphasized either as an exclusive or indeed as a principal factor (...)'.

However,

> 'For specialist wholesalers and retailers the desire to maintain a certain price level, which corresponds to the desire to preserve, in the interests of consumers, the possibility of the continued existence of this channel of distribution in conjunction with new methods of distribution (...) forms one of the objectives which may be pursued without necessarily falling under the prohibition contained in Article 85(1) (...)'.

The Court clarified in a later ruling on the same subject, *Metro II,* that some limitation on price competition is to be regarded as inherent in any selective distribution system, because the prices applied by specialist dealers necessarily remain within a much narrower margin than would be expected if there were competition between specialist dealers and non-specialist dealers[50]. It stated that that limitation was counterbalanced by competition as regards the quality of the services supplied to customers, which was not normally possible in the absence of an adequate profit margin covering the higher costs entailed by such services.

Whereas, in principle, selective distribution systems were not prohibited under Article 85(1), the Court indicated that there could be circumstances in which Article 85(1) would apply. As it explained in *Metro II*,

[50] Case 75/84 *Metro SB Großmärkte GmbH & Co KG v Commission* of 22 October 1986, (1986) ECR 3021. This case concerns an appeal by Metro against the Commission decision to extend the exemption of SABA's selective distribution system (the one which was the subject of the *Metro I* decision). Metro was of the opinion that this extension was not justified, in view of the changing market situation.

> 'there may nevertheless be a restriction or elimination of competition where the existence of a certain number of such systems does not leave any room for other forms of distribution (...) or results in a rigidity in price structure which is not counterbalanced by other aspects of competition between products of the same brands and by the existence of effective competition between different brands'.

As to the role of economic analysis under Article 85(1), the *Metro* judgment of the Court was very important. Not only did it imply a departure from the Commission's rigid interpretation of the concept of 'restriction of competition', it also carried a number of pertinent economic reasonings to determine whether or not an agreement could infringe Article 85(1). For example, the explicit reference to the need for specialist wholesalers and retailers to maintain a certain price level is a clear reference to the various 'free rider' arguments justifying vertical restraints. As we have seen in Chapter 2, when several retailers sell the same product, every retailer will benefit from it when other retailers make efforts to advertise the product or promote it by giving pre-sale advice. In such cases the incentives for retailers to make these kinds of effort may be too low (from the efficiency viewpoint), given that competitors who do not make these expenses can accordingly compete with sharp prices, while at the same time benefiting from the efforts made by other retailers. Selective distribution systems, as they limit competition in terms of prices and distribution format, may have the effect of restoring such incentives.

Similarly, the Court refers to the possibility of too much 'rigidity' in prices which may result from the co-existence of a number of selective distribution systems. As described in Chapter 2, the limitation of intrabrand competition by means of exclusive or selective distribution systems can be used by established producers as a strategic instrument in order to make competition among them 'softer'. In these forms of distribution systems, producers are facing a sales volume that reacts less strongly to rises in their own producer prices as dealers with some market power are inclined to partially absorb such rises; as a result final demand will drop by less than otherwise might have been the case. This may make producers feel more inclined to carry through price rises, amounting to a 'softening of competition'. The Court, in its reference to 'price rigidity' in the *Metro* judgments, does appear to recognise such an aspect.

The rule of reason was further developed and extended in various later judgments. The case *Nungesser* of 1982 concerned an exclusive licensing agreement between a French research institute and a German distributor for a new maize seed[51]. The Commission had found that the agreement infringed Article 85(1) on the following grounds. First, the Commission held that by licensing a single undertaking in a given territory the licensor would deprive himself of the ability to issue licences to other undertakings in the same territory. Further, by committing not to produce or market the product himself in the territory covered by the contract the licensor would likewise eliminate himself as a supplier in that territory. Finally, the fact that third parties would not be able to import the seed under licence from other Community countries into Germany (or export from Germany to other Community countries) would lead to market sharing.

The Court again applied a rule of reason in this case. It considered that exclusive licences can have an objective compatible with Article 85(1) and that, by consequence, the restrictions on economic freedom necessary for such licenses to be realized and maintained are not caught by Article 85(1):

> 'In fact, in the case of a licence of breeders' rights over hybrid maize seeds newly developed in one Member State, an undertaking established in another Member State which was not certain that it would not encounter competition from other licensees for the territory granted to it, or from the owner of the right himself, might be deterred from accepting the risk of cultivating and marketing that product; such a result would be damaging to the dissemination of a new technology and would prejudice competition in the Community between the new product and similar existing products'.

Notwithstanding this reasoning, the Court did stress that exclusive licence agreements could only escape the prohibition of Article 85(1) to the extent that they do not prevent parallel imports (with reference to the case *Consten and Grundig*). Therefore, the European 'rule of reason', while allowing for economic arguments to play a role in Article 85(1), had its limits.

[51] Case 258/78 *L.C. Nungesser KG and Kurt Eisele v Commission*, judgment of 8 June 1982, (1982) ECR 2015.

In the context of vertical agreements, the judgment *Pronuptia* of 1986 is of further interest[52]. This judgment, a preliminary ruling, was the first judgment about the validity of franchise networks. A German court had asked the European Court a few questions of law that arose in the context of proceedings between Pronuptia de Paris, a French company supplying wedding dresses and related articles, and one of its franchisees in Germany. Under the contested franchise contract, the franchisor granted the franchisee the exclusive right to use the Pronuptia trade-mark in respect of a specific territory, undertook not to open any other Pronuptia shops in the territory (or to provide goods or services to third parties) and to give the franchisee various forms of commercial and technical assistance. In turn, the franchisee, who was to remain sole proprietor of the business and take all the risks, would operate under the Pronuptia trademark and make the sale of bridal fashions its main purpose; sell the goods only in the premises specified in the contract; purchase most of its requirements for wedding dresses from the franchisor and refrain from opening a similar business or transferring know-how to third parties.

The Court took a fairly favourable view on franchise systems as such and held that certain provisions necessary to their establishment do not fall under Article 85(1). A franchise system was considered to be a means for a company, which had established itself in one market and thus developed certain business methods, to expand into other markets and to derive financial benefit from its expertise. At the same time, the system would give traders, who did not have the necessary experience, access to methods which they could not have learned without considerable effort and would allow them to benefit from the reputation of the franchisor's business name. Therefore, the Court held that a franchise system does not in itself interfere with competition.

Having established the principle, the Court was of the opinion that in order for a franchising system to work, two conditions must be met.

> 'First, the franchisor must be able to communicate his know-how to the franchisees and provide them with the necessary assistance in order to enable them to apply his methods, without running the risk that that know-how and assistance might benefit competitors, even indirectly. It follows that provisions which are essential in order to avoid that risk do not constitute

[52] Case 161/84 *Pronuptia de Paris GmbH v Pronuptia de Paris Irmgard Schillgallis*, judgment of 28 January 1986, (1986) ECR 353.

> restrictions on competition for the purposes of Article 85(1)'.

This applied to the clause prohibiting the franchisee from opening a shop of a similar nature (under a different name) and from transferring know-how to third parties.

> 'Secondly, the franchisor must be able to take the measures necessary for maintaining the identity and reputation of the network bearing his business name or symbol. It follows that provisions which establish the means of control necessary for that purpose do not constitute restrictions on competition for the purposes of Article 85(1)'.

This applied to the franchisee's obligation to apply the business methods developed by the franchisor and to use the know-how provided. Furthermore, as it would be impractical to lay down objective quality specifications for the distribution of fashion articles, the provision obliging the franchisee to sell only products supplied by the franchisor (or by suppliers selected by him) could be considered necessary for the protection of the network's reputation as well[53].

On the other hand, the Court emphasized that, far from being necessary for the protection of know-how or the maintenance of the network's identity and reputation, certain provisions are caught by Article 85(1) as they restrict competition between the members of the network, especially those provisions which share markets between franchisees or which prevent franchisees from engaging in price competition with each other. The Court particularly referred to the provision which obliges the franchisee to sell the goods in question only in the premises specified in the contract and to the exclusivity granted by the franchisor. According to the Court, these provisions restricted competition within the network as they prohibited the franchisee and the franchisor from opening other Pronuptia shops. In this respect, it recalled that

> 'As is clear from the judgment of 13 July 1966 (joined cases 56 and 58/64 *Consten and Grundig v Commission* (1966) ECR 299), a restriction of that kind constitutes a limitation of competition for the purposes of Article 85(1) if it concerns a business name or symbol which is already well-known. It is of course

[53] One may note that the particular justification given for the purchasing obligation very much fits into the incomplete contracts framework (cf. Chapter 2).

> possible that a prospective franchisee would not take the risk of becoming part of the chain, investing his own money, paying a relatively high entry fee and undertaking to pay a substantial annual royalty, unless he could hope, thanks to a degree of protection against competition on the part of the franchisor and other franchisees, that his business would be profitable. That consideration, however, is relevant only to an examination of the agreement in the light of the conditions laid down in Article 85(3)'.

Therefore, while the Court applied the rule of reason to a number of apparently 'restrictive' clauses in the franchise agreement, it did not go so far as to allow a clear territorial restriction of intrabrand competition under Article 85(1). One may note of course the qualifier 'if it concerns a business name or symbol which is already well-known', but this appears to hint at the requirement of appreciability, rather than an issue of market power at the interbrand level (Fasquelle, 1993).

More recently, in December 1994, the rule of reason was applied in the context of a buying cooperative[54]. This case concerned the statutes of a Danish cooperative association of distributors of agricultural basic products which said that a member of the cooperative was not allowed to be a member of another buying cooperative at the same time. The Commission had thought that this condition restricted competition (as the statutes prevented the members from freely obtaining supplies elsewhere), but not in an appreciable way. With it, the clause did not fall under Article 85(1). The Court, too, held the latter opinion, but for a different reason. According to the Court, in a market where product prices vary according to volume of orders, the activities of cooperative producing associations may, depending on the size of their membership, constitute a significant counterweight to the contractual power of large producers and make way for more effective competition. Therefore, while acknowledging that the clauses preventing the members from obtaining supplies elsewhere might have adverse effects on competition, the Court held that in so far the clauses are necessary for the proper functioning of the cooperative and for an effective bargaining position in relation to producers, they did not fall within the prohibition of 85(1). Another application of the rule of reason.

[54] Case C-250/92 *Gøttrup-Klim Grovvareforening and others v Dansk Landbrugs Grovvareselskab AmbA (DLG)*, judgment of 15 December1994, (1994) ECR I-5671.

The rule of reason can be seen as another important track in the jurisprudence of the Court which, at first sight, could be called well founded in economic respect. After all, the Court clearly states that the actual or intended effects of an agreement should be analyzed in the light of the market conditions with a view to establishing whether the agreement accords with Article 85(1). However, the rule of reason as applied by the European Court appears still a long way from the rule of reason as it is applied in the United States, in the sense of a purely competition based analysis (with a view to establishing whether an agreement or practice promotes or reduces economic efficiency[55]).

First of all, the Court has expressed its continued sensitivity towards agreements involving absolute territorial protection - also in the realm of the rule of reason - whereas there is no reason to assume that absolute territorial protection is always bad for competition or welfare. In Europe, however, the rule of reason has been placed in the context of 'workable competition', i.e. 'the degree of competition necessary to ensure the observance of the basic requirements and the attainment of the objectives of the EC Treaty, in particular the creation of a single market achieving conditions similar to those of a domestic market'. The Court's sensitivity clearly shows in its judgments on licensing agreements (*Nungesser*) and franchising agreements (*Pronuptia*) but also on selective distribution systems (*Metro*). For the latter type of distribution system, the Court has held that objective restrictions which are necessary for the system to be realized and maintained, such as qualitative criteria to appoint resellers and an obligation on resellers not to sell to non-approved resellers, are no restrictions of competition in the meaning of Article 85(1). By contrast, quantitative criteria relating to the number of eligible resellers and restrictions on resellers supplying other appointed resellers are considered restrictive of competition. Woods and Filipponi (1999) explain this choice by pointing out that arbitrage by network dealers is the only feasible mechanism to ensure parallel trade in a selective distribution network. The tighter the control which a supplier has over its distribution network the greater the potential to block parallel trade between network members[56]

Secondly, as Faull and Nikpay (1999) point out, the practical usefulness of the

[55] In the efficiency test as applied in the US, the emphasis lies on the effects on price and output levels (allocative efficiency). Even when companies achieve considerable cost savings (productive efficiency), the question remains whether some of these savings are passed on to consumers.

[56] While a qualitative system could be used for the same purpose, in theory the higher number of potential network dealers would limit the control exercised over individual network dealers and the potential for isolation of national markets.

European rule of reason is limited by the insistence of the Court on analysing the restrictive clauses in agreements falling outside Article 85(1) to ascertain whether they are *necessary* to secure the realisation or implementation of a lawful agreement[57]. In this sense, the European rule of reason is still very much an approach of *ancillary restraints*, restraints which are directly related to a lawful agreement and objectively necessary for its existence, but which remain subordinate in importance to the main object of the agreement[58]. Even though such an appraisal may be partly based on economic issues as well (such as the free rider problem), it differs from a true competition analysis weighing pro- and anticompetitive effects. In particular, as Gonzalez-Diaz (1995) points out, many restrictive clauses of the type discussed in this section are restraints which are necessary for the full preservation or full transfer of value in a transaction (the value embodied in the brand image of the supplier, intellectual property rights, know-how, etc.). The full preservation or transfer of such value often logically entails certain obligations on the part of distributors. Such restrictions, though based on *economic* considerations, do not require a *market* analysis to qualify for clearance under Article 85(1)[59] [60]. Finally, as indicated by Goyder (1993), the European rule of reason would not apply for clauses which are not necessary for the existence or implementation of agreements even when these are, in the end, not restricting competition. These elements of the European rule of reason appear, however, less present in the more recent approach taken by the Court in the recent cases on foreclosure issues. These will be discussed in the following subsection.

[57] The name 'European rule of reason' is used here for consistency reasons. Faull and Nikpay (1999) do not use this name.

[58] Commission Notice concerning the assessment of co-operative joint ventures pursuant to Article 85 of the EEC Treaty, 1993 Official Journal C43/2.

[59] There are some clauses that are borderline, such as the commitment not to resell to unauthorised dealers in the case of selective distribution. In principle, such a restriction automatically qualifies as ancillary. However, the restriction falls under Article 85(1) where the existence of a certain number of selective distribution systems does not leave any room for other forms of distribution or results in too much rigidity in prices. Such an appraisal clearly requires a market analysis.

[60] Along the same lines, Gonzalez-Diaz (1995) argues that the distinction between qualitative and quantitative restrictions in the context of selective distribution systems is perhaps best explained by the fact that the former do not normally require a market analysis to qualify for clearance under Article 85(1), whereas the latter do, since it is difficult to establish from the outset whether or not quantitative restrictions are objectively necessary for the realisation or implementation of a selective distribution system.

3.4.4 The cumulative effect doctrine

Of particular interest for the role of economic analysis within the framework of Article 85(1) are the judgments of the European Court with regard to the foreclosure problems in exclusive dealing agreements[61]. With respect to such agreements there is some concern that the distributors present in a market should not be tied too much to the producers (in number or duration), in order to keep it possible for new producers to find distributors willing to carry their product. The Court expressed a clear opinion on the issue of foreclosure in the case *Delimitis* of 1991[62]. In this case, the Court was asked to give a preliminary ruling concerning a contract dispute between a pub tenant, Mr. Delimitis, and the German brewer Henninger Bräu. Building on the judgment *Brasserie de Haecht* of the 1960s, the Court indicated

> 'If an examination of all similar contracts entered into on the relevant market and the other factors relevant to the economic and legal context in which the contract must be examined shows that those agreements do not have the cumulative effect of denying access to that market to new national and foreign competitors, the individual agreements comprising the bundle of agreements cannot be held to restrict competition within the meaning of Article 85(1) of the Treaty. They do not, therefore, fall under the prohibition laid down in that provision'.

If, on the other hand, such examination would reveal that it is difficult to gain access to the relevant market, then it is necessary to assess the extent to which the agreements entered into by the brewery in question contribute to the cumulative effect produced in that respect by the totality of similar contracts found on that market.

> 'Under the Community rules on competition, responsibility for such an effect of closing off the market must be attributed to the

[61] The terms 'exclusive dealing' and 'exclusive purchasing' will be used interchangeably, even though they are, strictly speaking, not synonymous (see also recital 10 of the Court's judgment in case *Delimitis* discussed further below). Under EC competition law, exclusive purchasing refers to an obligation to purchase the goods that are traded under the contract from the supplier only. In economics, exclusive dealing refers to an obligation not to distribute goods which are competing with those of the supplier in question. In practice, the two will often go hand in hand (but not for selective distribution systems).

[62] Case C-234/89 *Delimitis S. v Henninger Bräu*, judgment of 28 February 1991, (1991) ECR 935.

> breweries which make an appreciable contribution thereto. Beer supply agreements entered into by breweries whose contribution to the cumulative effect is insignificant do not therefore fall under the prohibition under Article 85(1)'.

In other words, an exclusive dealing contract only falls within 85(1) if the cumulative effect of all similar agreements together results in the situation that access to a market is limited and the agreements of the producer in question contribute to this to some degree. In other cases, the agreement is not caught by Article 85(1). What is noteworthy, is that in this respect no specific reference was made to the exclusivity clauses being necessary to attain certain goals to be achieved[63].

The judgment is also interesting in other respects, such as the emphasis the Court placed on the need to determine the relevant market and to fully assess whether the existence of several beer supply agreements impedes access to the market. The Court did not just refer to elements such as the existing proportion of tied pubs and the length and nature of the exclusive agreements, but equally to the question whether there are real concrete possibilities for a new competitor to penetrate into the bundle of contracts by acquiring a brewery already established on the market (together with its network of sales outlets) or to circumvent the bundle of contracts by opening new pubs.

As *Delimitis* made clear, an exclusive dealing contract can be caught by Article 85(1) if the cumulative effect of all similar agreements together results in the situation that access to a market is foreclosed and the agreements of the producer in question contribute to this to some degree. Some uncertainty remained about the question how the cumulative effect doctrine interacted with the requirement of appreciability. This is illustrated by the approach taken by Commission in the cases *Langnese-Iglo* and *Schöller* of 1995, on exclusive dealing contracts for ice cream in Germany[64]. The Commission had maintained that it is only where the network of similar agreements concluded by the producer concerned does not by itself have any appreciable effect that the rules developed by the Court in *Delimitis* require an examination of the

[63] Rather, by way of introduction, the clauses were described as an inherent feature of the co-operation between reseller and supplier, based on their shared interest in promoting sales of the products.

[64] Case T-7/93 *Langnese Iglo GmbH v Commission*, judgment of 8 June 1995, (1995) ECR II-1533. Case T-9/93 *Schöller Lebensmittel GmbH & Co v Commission*, judgment of 8 June 1995, (1995) ECR II-1615.

cumulative effects of parallel networks[65]. This particular way of interpreting *Delimitis* was, however, not accepted by the Court. On the other hand, also on the issue of appreciability, the Court specified that if a bundle of contracts of a producer is found to restrict competition, this implies that all individual contracts with its distributors are caught by Article 85(1), the small ones as well as the large ones.

3.4.5 Towards a real competition balance ?

In the jurisprudence of the Court we can find (at least) three important tracks which, at first sight, could be called well founded in economic respect: the requirement of appreciability, the European 'rule of reason' and the approach towards foreclosure. So, possible starting points for further economic analysis under Article 85(1) can be considered to be present in the jurisprudence. On the other hand, it is difficult to maintain that these three approaches are amounting to the rule of reason as it is applied in the United States, in the sense of a purely competition based analysis, weighing positive and negative effects on competition. As explained above, the requirement of appreciability has been intended to separate agreements with a Community interest from those which had better be left to the member states. The European rule of reason is still very much an approach of ancillary restraints, restraints necessary for the implementation of a lawful agreement. Finally, the cumulative effect doctrine is more about establishing that certain agreements are not restricting competition (e.g. because they leave sufficient market access), rather than an approach open to a true balancing of possible positive and negative effects.

However, it is fair to say that in a growing number of cases, the European Courts have looked beyond the negative elements of agreements in their analysis under Article 85(1) and have sought to determine what the actual or potential economic impact on the market could be. In this respect, a final comment deserves to be devoted to the recent judgment *European Night Services*, in which the Court appears to have

[65]Given that both Langnese's and Schöller's market shares surpassed the Commission's benchmark of appreciability set out in its *de minimis* Notice, the Commission had held that the agreements were caught by Article 85(1). Hawk (1995) characterises the Commission's explanation of the Delimitis judgment in these cases as another example of the Commission being reluctant in taking on board indications given by the Court. Also Korah (1998) observes that the Commission has considered a number of agreements to fall within Article 85(1), of which the Court had ruled that such agreements did not restrict competition.

further refined the scope of application of Article 85(1)[66] [67]. This case did not concern a vertical agreement as such, but a joint venture agreement between European rail companies and some related agreements (for the provision of the necessary infrastructure and services), which the Commission had exempted under Article 85(3) under fairly strict conditions. In response, the parties to the agreement brought an appeal against this decision, claiming among other things that the agreements did not infringe Article 85(1) in the first place.

As regards the overall assessment of the agreement, the parties were of the opinion that 'the Court of Justice has consistently held that the pro-competitive effects of an agreement must be weighed up against its anti-competitive effects', referring specifically to the cases *STM v. Maschinenbau Ulm*, *Consten and Grundig*, *Metro I*, *Nungesser*, *Pronuptia* and *Delimitis*. They added: 'If the pro-competitive effects outweigh the anti-competitive effects and the latter are necessary in order to implement the agreement, then the agreement cannot be regarded as having as its object or effect the prevention, restriction or distortion of competition within the common market within the meaning of Article 85(1) of the Treaty'. The UK government, in intervention, submitted that in applying Article 85(1) the Commission had failed to take account of the economic context and, in particular, of the state of competition that would exist in the absence of the agreements. These would not restrict competition because they were designed to facilitate, and were necessary for, the introduction of a service which is not currently operating and which none of the parties could reasonably be expected to introduce by itself.

The Commission submitted that, while the analysis of an agreement must take account of its economic context, it did not follow that the rule of reason should be used. According to the Commission that conclusion was not negated by the Court's judgment in the Danish buying cooperative case, which only concerned the validity of ancillary restrictions in the specific context of cooperative organisations and could not, therefore, be regarded as the expression of a general principle. Consequently, the Commission held that balancing the competitive benefits and harms of an agreement should take place under Article 85(3) and not under Article 85(1).

In its wording, however, the Court appears to have left open the possibility for such a 'rule of reason' approach where restrictions by effect are involved (Faull and Nikpay,

[66] Joined cases T-374/94, T-375/95, T-384/94, T-388/94 *European Night Services (ENS) v Commission*, judgment of 15 September 1998, (1998) ECR II-3141.

[67] Cf. Woods and Filipponi (1999) and Faull and Nikpay (1999).

1999). It did so, by expressly excluding the weighing up of the pro-competitive effects of an agreement against its anti-competitive effects *only* where restrictions by object are involved:

> '(...) it must be borne in mind that in assessing an agreement under Article 85(1) of the Treaty, account should be taken of the actual conditions in which it functions, in particular the economic context in which the undertakings operate, the products or services covered by the agreement and the actual structure of the market concerned (...), unless it is an agreement containing obvious restrictions of competition such as price-fixing, market-sharing or the control of outlets (...). In the latter case, such restrictions may be weighed against their claimed pro-competitive effects only in the context of Article 85(3) of the Treaty, with a view to granting an exemption from the prohibition in Article 85(1)'

3.5 Towards a new enforcement system

Throughout the years, there have been regular calls for a more economics based analysis of the concept 'restriction of competition' under Article 85(1), among others from commentators such as Korah (1981), Waelbroeck (1987) and Hawk (1989, 1995) [68]. They expressed criticism on the Commission's formalistic approach and on the Commission's reluctance to take on board the signals given by the European Court of Justice in its various judgments. As Hawk (1995) notes, the Court's use of the rule of reason and the requirement of appreciability could be used as *doctrinal vehicles* for a more economics based analysis under Article 85(1).

Opponents to this view, in particular from the side of the Commission, have indicated that the emphasis of economic analysis lies on Article 85(3) rather than Article 85(1). The Commission itself has traditionally held that 'Article 85 provides an appropriate legal framework for a balanced assessment, recognising the distinction between anticompetitive and procompetitive effects. Article 85(1) covers those agreements which appreciably restrict or distort competition. Article 85(3) allows for the exemp-

[68] Cf. Van Gerven e.a. (1997).

tion of such agreements provided that they have sufficient efficiency benefits'[69]. In this respect, the structure of Article 85 is 'such as to prevent greater use of the rule of reason being made: if more systematic use were made under Article 85(1) of an analysis of the pro- and anticompetitive aspects of a restrictive agreement, Article 85(3) would be cast aside, whereas any such change could be made only through revision of the Treaty'[70]. Furthermore, as Fasquelle (1993) points out, a further development of the rule of reason under Article 85(1) would be contradictory to the spirit of Regulation 17, which gave to the Commission the exclusive power to make an economic assessment under the criteria of Article 85(3).

In a way, it does not really matter a lot where the economic analysis takes place, as long as it does take place and in a way not too much burdening the business community. As described in the introduction to this chapter, precisely this appears to haven been the main thrust in the criticism on the Commission's policy in the past. The notification system together with the strict interpretation of Article 85(1) led to a considerable administrative burden resulting from the fact that too many agreements, also utterly innocent ones, ended up in the net of the European competition procedures. At the same time, the traditional group exemptions were very closely defined and, hence, little flexible. By this, the formation of distribution systems adapted to the needs of the companies involved was not particularly encouraged. Strangely enough, the block exemptions equally applied to agreements concluded by companies with very strong market positions. In this sense, the Commission's policy was not very efficient in making a distinction between the competition enhancing effects of vertical agreements and the effects restricting competition.

Whether it has been under the influence of the criticisms made, the indications of the Court as regards the necessity of economic analysis under Article 81/85(1) or otherwise, the Commission has changed opinion on its course of policy in the meantime. In 1997, it issued a Green Paper (a consultation document) on EC competition policy in the domain of vertical restraints[71]. In this Green Paper, a prominent place

[69]Commission Notice - Guidelines on Vertical Restraints, para 1 (Official Journal C 291, 13 October 2000, p.21; old numbering for Article 81 retained). Similarly, one can read in para 112 of the Guidelines that a vertical restraint falls under Article 81(1) when it is 'likely to have an appreciable negative effect on competition'. According to Wils (2000), Article 85(3) is nothing but the codification of the (American) rule of reason in European competition law. This view is, however, not uncontested (cf. Wesseling, 1999).

[70]European Commission: White Paper on modernisation of the rules implementing Articles 85 and 86 of the EC treaty, 28 April 1999. See also Marenco (1999).

[71]Green Paper on Vertical Restraints in EC Competition Policy, 22 January 1997, COM(96) 721

was given to an overview of the main insights from the economic science as regards the effects of vertical agreements. Main element: it depends on the circumstances of the case in question, particularly on the presence or absence of sufficient interbrand competition, whether a vertical agreement is harmful or not. In a follow-up, the Commission concluded that it is only when interbrand competition is weak and market power exists, that it becomes important to control vertical agreements[72].

In 1999, the Commission put these conclusions into effect by issuing a new Block Exemption Regulation (BER)[73]. In principle, this Regulation applies to all distribution agreements (including selective distribution) and thereby seeks to avoid a bias towards any of them[74]. According to the Commission, the BER is meant to create a 'safe harbour' for vertical agreements concerning the sale of goods and services which are concluded by companies with less than 30 percent market share, except for a limited number of so-called 'hard-core' restraints such as restrictions on parallel imports and resale price maintenance. Additionally, even when the market share threshold is not exceeded, there is a number of conditions which must be met if the agreements are to benefit from the block exemption. These relate principally to exclusive dealing contracts (they should be limited to 5 years), post-term non-compete obligations and to 'too prescriptive' selective distribution agreements[75] [76].

final.

[72]Communication from the Commission on the application of the Community competition rules to vertical restraints - Follow-up to the Green Paper on vertical restraints. COM(98) 544.

[73]Commission Regulation (EC) No. 2790/99 on the application of Article 81(3) of the Treaty to categories of vertical agreements and concerted practices, OJ L 336, 29 December 1999, p. 21.

[74]There are a few exceptions. For the moment, motor vehicle distribution and agency agreements are not covered by the Regulation. Furthermore, while it is 'one Regulation for all', there are specific provisions for the specific types of distribution agreements. For example, in selective distribution agreements obligations to sell or not to sell competitors' products imposed by a supplier on its distributors are not exempted by the Regulation.

[75]The B.E. Regulation contains the condition that a supplier may not impose an obligation on the appointed dealers to sell or not to sell specified brands of competing suppliers. This condition is meant to avoid horizontal collusion through the creation of a selective club of brands from the leading suppliers as well as foreclosure of market access.

[76]In addition to these conditions, the Commission and, in some cases, the national competition authorities are empowered to withdraw the benefit of the block exemption for a particular agreement when the cumulative effect of parallel networks of similar vertical agreements practiced by competitors leads to foreclosure of the market or a restriction in competition ('withdrawal' of the block exemption). In addition, if such parallel networks cover more than 50 percent of a relevant market, the Commission can exclude them from the scope of the block exemption. In this case, the measure will be addressed not to the individual companies but to all undertakings whose agreements come within the scope of application of the block exemption ('disapplication' of the block exemption regulation). See Kmiecik (2000) for a detailed overview.

Above the threshold of 30% or if the stipulated conditions are not met, vertical agreements will not be presumed to be illegal but may need individual exemption. For this purpose, the BER is accompanied by a detailed set of Guidelines which set out the policy that the Commission will adopt in applying the competition rules in these circumstances[77]. These Guidelines are, in general, reflecting the types of arguments described in Chapter 2 and a recognition of the fact that it is only when market power exists that vertical agreements can have negative effects on competition. By contrast, if an agreement contains a 'hard-core' restriction, it is almost sure to be considered illegal (provided that the agreement is not *de minimis*).

In order not to punish firms which make unintentional mistakes in the assessment of their market positions, legislative changes have been instituted so that vertical agreements may be exempted retrospectively from the prohibition in Article 81(1)[78]. The Commission hopes that these measures will reduce the number of notifications, since the failure to notify an agreement containing vertical restraints upfront will not mean that the agreement is unenforceable, if an exemption can be granted[79].

Starting from the observation that vertical agreements can only be harmful when interbrand competition is limited, these proposals are a clear improvement. Although, of course, the criterion of market share remains a somewhat imperfect indicator for the existence of sufficient interbrand competition. This is especially so, given that - as Nazerali and Cowan (1999) describe it - working out the relevant market is not an exact science. The Commission has traditionally tended to define the relevant market rather narrowly, both in the product dimension and in the geographic dimension (Bishop and Walker, 1999). For the new system to work it is essential that the Commission adopts a clear and reliable approach to market definition. Only in this way will companies be able to apply the block exemption themselves in the first instance and not be facing excessive legal uncertainty.

On a more general level, Griffiths (2000) is of the opinion that despite the rhetoric

[77] Commission Notice - Guidelines on Vertical Restraints, Official Journal C 291, 13 October 2000.

[78] Council Regulation (EC) No 1216/1999 amending Regulation No 17 (Official Journal C 291). In effect since 18 June 1999.

[79] Whether this hope is justified remains to be seen. After all, in the event that the Commission would refuse an exemption or make it conditional upon changes being made to the agreement, this would oblige the contracting parties to renegotiate their agreement. This may be a very delicate matter, since the bargaining positions may have changed quite substantially since the moment the original contract was concluded (cf. the literature on asset specificity and incomplete contracts discussed in Chapter 2). From this perspective, companies would be advised to notify where there is any doubt as to the compatibility with competition rules.

of the Commission in relation to the abandonment of a formalistic approach, the Regulation is still plagued by an ignorance of economics in that the black list of prohibited clauses is too long and that there is an over-emphasis on the importance of market share evaluation. In this respect, he characterises the BER as no more significant than an extension of the *de minimis* Notice with the market share threshold being raised to 30%. He refers in particular to the fact that the European Courts have indicated that also agreements of the type excluded from the Regulation need not fall under Article 81(1) in the first place.

It does appear that the question whether European competition policy will remain formalistic depends more on the way the Commission is going to deal with agreements which do not fall under the BER, rather than on the wording of the Regulation itself. The Guidelines themselves are not always completely reassuring in this respect. For example, 'in case the market share of the manufacturer is above 30% (...) foreclosure of price discounters may make the selective distribution system difficult to exempt (...)'[80]. In view of the fact that the Commission has a history of treating agreements which are over and above the *de minimis* threshold of 10% as almost automatically restricting competition (cf. the *Langnese-Iglo* and *Schöller* cases), Griffiths' scepticism is not unfounded.

Finally, from a pure competition perspective, the continued strict treatment of absolute territorial protection remains a downside, even though there are some improvements[81]. Not that absolute territorial protection is necessarily a good thing, but a per se prohibition is not very appropriate from the efficiency viewpoint. But then, as discussed in Section 3.3, the EC rules on competition are to be seen in the overall context of the Treaty with its historic emphasis on bringing down conventional impediments to trade between member states. This justification does however not exist for resale price maintenance, which also continues to be treated as per se prohibited (unless *de minimis*).

[80] Commission Notice - Guidelines on Vertical Restraints, para 190 (Official Journal C 291, 13 October 2000).

[81] By way of exception, in the case of the introduction of a product in a new geographic market, a fairly high degree of territorial protection is allowed for two years: a supplier can oblige its existing resellers not to sell to unofficial resellers in the new geographic market. Absolute territorial protection is not allowed, however: a supplier cannot prevent its resellers from making passive sales (i.e. unsollicited sales) to end-users from the new geographic market. Of considerable importance is the fact that internet sales are normally considered by the Commission to be passive sales.

3.5.1 The proposal for a new 'Regulation 17'

In the same year as in which it adopted the new Block Exemption Regulation, in 1999, the Commission also embarked upon a different, far reaching operation, namely to reform Regulation 17 (the enforcement system of Article 81) altogether[82]. In an attempt to increase the role of national competition authorities and national courts in enforcing Article 81, the Commission proposes to replace the current centralised notification and exemption system by one in which not only the Commission but also the national authorities and courts will be able to apply Article 81 *as a whole*. To this end, the Commission submitted on 27 September 2000 a proposal for a new Regulation to the Council of Ministers[83].

The fact that national competition authorities and courts will be able to apply Article 81 as a whole does not mean that they will be able to grant exemptions. As a matter of fact, exemptions as we know them will cease to exist, as will notifications for that matter (Wils, 2000). The new competition law system will be a prohibition system with a directly applicable exception rule whenever the criteria of Article 81(3) are met. This means that in all proceedings in which the relevant authorities[84] are called upon to apply the prohibition rule of Article 81(1), they are called upon to apply Article 81(3) as well, if at all necessary. This contrasts with the current authorisation system which is based on the principle that the prohibition contained in Article 81(1) can be lifted only by a (preceding) act of a public authority empowered to do so (under Regulation 17, this is the Commission). In other words, Article 81 is to be applied as a single legal norm, establishing a principle of prohibition unless certain conditions are met.

The main reason why the Commission insists on this reform is that it feels that a single institution - the Commission - cannot on its own ensure that the EC competition rules are complied with. The Commission expects that a decentralisation of law enforcement leads to a more effective enforcement system and a dissemination of knowledge of the EC competition rules towards a level closer to the business

[82]The Commission initiated this policy reform by issuing the so-called White Paper on modernisation of the rules implementing Articles 85 and 86 of the EC Treaty (Commission Programme No. 99/027).

[83]Proposal for a Council Regulation on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty, COM(2000) 582 final.

[84]Pursuant to the new Regulation, these are the Commission, the national competition authorities and the national courts. Article 81(1) was already directly applicable by the national courts under Regulation 17 (cf. Section 3.2).

community and the consumers. For example, it is hoped that the reform will foster meaningful litigation activity before the national courts. Furthermore, the abolition of the current notification and exemption system is supposed to enable the Commission to free up resources and to start targetting the most serious restrictions such as price fixing cartels or market sharing agreements.

¿From a legal perspective, the Commission explains this policy step by observing that the 'current division between paragraph 1 and paragraph 3 in implementing Article 81 is artificial and runs counter to the integral nature of Article 81, which requires economic analysis of the overall impact of restrictive practices'[85]. Effectively, it emerges clearly from the development of the jurisprudence of the Court of Justice as described in Section 3.4 that the distinction between agreements not restricting competition in the sense of Article 81(1) and agreements exemptable under Article 81(3) has become increasingly blurred. From a practical perspective, the widening of the scope of the block exemption system to the majority of vertical agreements concluded between parties with less than 30% market share has put the distinction further into the background. After all, it is only when the block exemption regulation does not apply, that the distinction becomes a live issue. In these respects, viewing Article 81 as a single legal norm is a logical and welcome step.

These policy initiatives are not uncontested. Especially the German competition authorities, the Bundeskartellamt and the Monopolkommission, have expressed some fierce criticisms[86]. These refer in the first place to the compatibility of the proposed reform with the EC Treaty itself, in that the wording of Article 81 would suggest that an agreement can only benefit from Article 81(3) following the authorisation by some authority. It goes beyond the scope of this chapter to comment on the debate concerning the original intentions of the authors of the EC Treaty, except to say that it is a most interesting one.

Further, opponents raise the question whether Article 81(3) is really suited for direct applicability, i.e. whether its criteria are sufficiently precise in order to have a direct effect. The Commission has indicated that 30 years of case practise and jurisprudence of the European Court of Justice provide sufficient guidance to take the proposed initiative. Leaving aside the question whether the jurisprudence is com-

[85] White Paper on modernisation of the rules implementing Articles 85 and 86 of the EC Treaty (Commission Programme No. 99/027), para 49; new numbering inserted.
[86] Cf. Ehlermann (2000).

pletely clear and consistent[87], the fact remains that one of the determining factors behind the interpretation of Article 81(1) has been the particular division of competences between the Commission and the member states and that exactly this division is now disappearing. Even in case the different sets of arguments which can be relied upon under the two paragraphs are clear[88], the distinction between Article 81(1) and 81(3) remains relevant, given that the resulting burden of proof differs. According to Article 2 of the proposed Regulation, the burden of proving an infringement of Article 81(1) in any national or Community proceedings rests on the party alleging the infringement. A party claiming the benefit of Article 81(3) shall bear the burden of proving that the conditions of that paragraph are fulfilled[89].

A decentralized application of Art. 81 is intended to bring an additional number of cases before the national courts. This is also the reason why the Commission has not opted for a partial decentralisation involving the national competition authorities only[90]. However, it is not to be taken for granted that each and every national court has sufficient expertise in and knowledge about competition cases. One may even raise the question whether the national courts themselves have the time, resources and interest in dealing with competition cases. Wolf (2000) refers in this respect to the proverbial 'judge in Palermo'. As a matter of fact, it is only very recently that in the Netherlands a specific competition authority was founded, among other reasons because it was felt that competition cases were typically not given much priority in the courts supposed to deal with them[91]. Apart from this, there is a certain risk that busy judges or 'economically illiterate' judges may choose to hang on to the clear-cut market share benchmarks of the BER (which will continue to exist as a way to declare certain agreements compatible with Article 81), bringing again a certain formalism in

[87]For example, Goyder (1993) indicates that referring to a 'rule' in the context of Article 85(1) leads to 'confusion rather than clarity'. Faull and Nikpay (1999) describe the definition given to ancillary restraints under European competition law as 'imprecise'.

[88]There is a continuing debate as to whether or not Article 81(3) provides room for policy objectives which are unrelated to competition or market integration. Cf. Wesseling (1999) and Wils (2000).

[89]As for its own policy, the Commission has stated that there is not a presumption of illegality outside the scope of the regulation, and that the burden of proof will lie on itself to demonstrate the contravention of Article 81(1). As Griffiths (2000) observes, if this demonstration occurs in a loose way similar to past practise, then the burden will not remain on the Commission for a long time.

[90]This could have been achieved by leaving the current enforcement system intact but sharing the power to apply Article 81(3) with the national competition authorities and not with the national courts.

[91]Netherlands Ministry of Economic Affairs, 'Nieuwe regels omtrent de economische mededinging (Mededingingswet)', proposal for a new competition law, May 1996. To be precise, the argument related to public law enforcement rather than civil law enforcement.

the application of the EC competition rules[92].

Another major concern is the need for a uniform, Community-wide application of the EC competition rules. With many national competition authorities and courts applying Article 81 as a whole, there is a certain risk of divergences occurring in their approach towards restrictive agreements. Several safeguards are envisioned to avoid such divergences from happening. First, as already noted, the notion of exemption will become very different. In the new system, no authority or court, European or national, will have the power to adopt positive decisions which immunize an agreement against attack by other competition authorities or other courts (Wils, 2000). For decisions requiring that an infringement be brought to an end, accepting commitments or withdrawing the benefit of a block exemption, national competition authorities are required to consult the Commission. The Commission can then take over the procedure by initiating proceedings itself. Further, the new Regulation foresees a right for the Commission (acting in the Community public interest) to make written or oral submissions in national proceedings. Finally, a number of provisions on information sharing and case attribution are to bring about a real 'network' of competition authorities.

Wolf (1999) makes two further criticisms which are worth mentioning. One of the main reasons to start this reform is to alleviate the administrative burden faced by companies which had to notify their agreements under Regulation 17 in order to qualify for an exemption. Wolf wonders whether one can really speak of alleviating a burden when companies must in the future go to law firms in order to obtain a competition law assessment of their agreements, because the Commission no longer gives binding statements. One might counter this argument by observing that the Commission did not issue many binding statements anyway, in view of the sheer number of notifications and the resulting backlog. In this respect, as Fasquelle (1993) notes, the Commission system had, to some extent, discredited itself.

Secondly, Wolf (1999) is worried by another aspect of enforcement, namely the possibility of fining companies who have enacted agreements in breach of Article 81. Under the old system, if companies enacted 'audacious' agreements without notifying

[92]Still, the Guidelines (if properly used) should loosen up things. Furthermore, it is comforting that the new regulation foresees a right for the national competition authorities and the Commission (acting in the Community public interest) to make submissions to national courts in written or oral form. Similarly, national courts will, as before, be able to request a preliminary ruling form the European Courts. This, however, will require resources from their side just the same.

them, there was a certain presumption of illegality in view of Regulation 17's clear obligation to notify agreements if they are to be exempted. With the new system, it will become more difficult to negate the parties' natural claim that they were living in the opinion that the agreement did not infringe Article 81[93]. Wolf submits that it will therefore become more difficult for the Commission to impose fines, leading to a possible reduction of adherence to competition law. Doubtlessly, much will depend on the way in which the Commission is going to deal with this issue and even more so on the question whether the European Courts will provide support for it.

3.6 Concluding remarks

At this very moment, many developments are taking place in the policy field towards vertical restraints. From the viewpoint of evaluating the pro- and anticompetitive effects of vertical agreements under Article 81, a number of good changes have been started up. On the substantive side, the new Block Exemption Regulation can be seen as increasing the role for economic analysis in the application of Article 81. The more flexible and permissive character of the new Regulation is to be welcomed. Nonetheless, it would appear that also the new policy has a certain procedural flavour, this time through the extensive use of market share thresholds. The fact that the European Commission decides to shape the changes in substance into a 'procedural jacket' is, for the purpose of legal certainty, far from wrong. The point is that the jacket should not be unnecessarily tight.

An important role in this respect is given to the determination of the relevant product markets and their geographic dimensions. After all, it is on the basis of this determination that the market shares are calculated and compared with the legal thresholds. The Commission has traditionally tended to define the relevant market rather narrowly, both in the product dimension and in the geographic dimension. In my opinion, if the adoption of the new economic approach towards vertical restraints is to have its full effect, it is not enough to change a Block Exemption Regulation. It is also necessary to be careful in determining the relevant product markets and geographic markets. This does require sufficient discipline form the side of the Commission.

[93] Wolf (1999) made this remark in the context of horizontal agreements, rather than vertical agreements. After all, the obligation to notify vertical agreements for exemption has already become less pressing in view of the fact that they can be exempted retrospectively (since 1999).

The Commission's proposal to share the enforcement of Article 81 with the member states is a good proposal. It is not very realistic to assume that one institution can on its own monitor a market with more than 350 million consumers. The fact that national courts will be called upon to apply not just Article 81(1), but also Article 81(3), will bring the EU competition rules a lot closer to a level where they are seen to exist. At the same time, also the substantive distinction between agreements not restricting competition in the sense of Article 81(1) and agreements exemptable under Article 81(3) had become increasingly blurred, in view of the jurisprudence of the Court of Justice. In these respects, viewing Article 81 as a single legal norm is a logical and welcome step.

There remains a practical problem, however, with the application of Article 81. The burden of proving an infringement of Article 81(1) rests on the party alleging the infringement. A party claiming the benefit of Article 81(3) shall bear the burden of proving that the conditions of that paragraph are fulfilled. As for its own policy, the Commission has stated that there is not a presumption of illegality outside the scope of the Block Exemption Regulation, and that the burden of proof will lie on itself to demonstrate the infringement of Article 81(1). Much will depend on the Commission's own style in market definition and the actual assessment of vertical restraints, whether this burden is at the requisite level.

This brings us back to the main theme of the chapter: the influence of the European Courts on the Commission's policy on vertical agreements in the course of the years. In case too much formalism were to develop in the Commision's new policy, there is still the European Court of Justice to turn to. Possibly that with arguments of an economic nature - of which we have seen many examples - things can be given the necessary flexibility. After all, as far as economic argumentation is concerned: the Court has often been responsive to it.

3.7 Appendix A: Article 81 of the EC Treaty

Article 81 of the EC Treaty (ex Article 85)

1.The following shall be prohibited as incompatible with the common market: all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the common market, and in particular those which:
(a) directly or indirectly fix purchase or selling prices or any other trading conditions;
(b) limit or control production, markets, technical development, or investment;
(c) share markets or sources of supply;
(d) apply dissimilar conditions to equivalent transactions with other trading parties, thereby placing them at a competitive disadvantage;
(e) make the conclusion of contracts subject to acceptance by the other parties of supplementary obligations which, by their nature or according to commercial usage, have no connection with the subject of such contracts.

2.Any agreements or decisions prohibited pursuant to this Article shall be automatically void.

3.The provisions of paragraph 1 may, however, be declared inapplicable in the case of:
- any agreement or category of agreements between undertakings;
- any decision or category of decisions by associations of undertakings;
- any concerted practice or category of concerted practices,

which contributes to improving the production or distribution of goods or to promoting technical or economic progress, while allowing consumers a fair share of the resulting benefit, and which does not:
(a) impose on the undertakings concerned restrictions which are not indispensable to the attainment of these objectives;
(b) afford such undertakings the possibility of eliminating competition in respect of a substantial part of the products in question.

3.8 Appendix B: the American 'rule of reason'

The rule of reason is a concept that takes a central position in American antitrust law. In American antitrust, the rule of reason entails an assessment of the net effect of the agreement on competition, implying a comparison between the pro-competitive and anti-competitive effects in the light of the prevailing market circumstances, without reference to any other elements having no bearing with competition (Van Gerven e.a., 1997).

The American rule of reason, as we know it today, is to be seen in the context of the general prohibition principle imposed by the Sherman Act of 1890. Section 1 of the Sherman Act states that 'every contract (...) in restraint of trade or commerce among the several States (...) is hereby declared to be illegal'. Already at an early stage, and fully in line with the principles of the anglo-saxon Common Law system[94], the Surpreme Court took the view that the qualification 'every' could impossibly mean that, without exception, every restraint of trade between contracting parties should be presumed a violation of Section 1 of the Sherman Act. Therefore, in *Addyston Pype* (1899), the Surpreme Court declared that restraints necessary for a commercial deal to come about ('ancillary restraints') are not a violation of Section 1 if this deal has a lawful (neutral) effect on the state of competition[95] [96]. Still, to allow any other

[94]The adoption of the Sherman Act, though being of great importance, was not by itself a shift in the approach towards restraints of trade (Fasquelle, 1993; Marenco, 1999). The restraints of trade doctrine had since long been applied by the local courts, in the tradition of the Common Law system. According to Senator Sherman: 'The purpose of this bill is to enable the courts of the United States to apply the same remedies against combinations which injuriously affect the interests of the United States that have been applied in the several States to protect local interest' (Congress Rec. 1890, Vol. 21, p. 2456). Neither was the initial application of the 'rule of reason' a change with the past. The restraints of trade doctrine and the 'rule of reason' go, in fact, much further back in history and have their origin in the English Common Law tribunals of the late Middle Ages (Van Gerven e.a., 1997). At the time, the English jurisdictions considered that the freedom to trade had to be protected and that contract clauses restricting the freedom of individuals were at odds with this principle. It was however impossible for the judges to apply this principle without any nuance. A distinction was therefore being made between contractual restrictions which were 'reasonable' and those which were not. Requirements for a contractual restriction to be 'reasonable' was that it made sense in the context of the agreement (e.g. that it protected the value of a business to be sold) and that there was a 'consideration' for the party (parties) taking up the restriction, i.e. something in return. In a sense, the rule of reason was a means to put into effect a general principle of prohibition.

[95]In *U.S. v Joint Traffic* (1898), the Supreme Court had indicated that the wording 'restraint of trade' should be understood as 'restraint of competition'.

[96]*Addyston Pype and Steel Co v. United States* 166 US 211 (1899) 'No conventional restraint of trade can be enforced unless the covenant embodying it is merely ancillary to the main purpose of a lawful contract and necessary to protect the covenantee in the enjoyment of the legitimate fruits of the contract, or protect him from the dangers of an unjust use of those fruits by the other party'.

restraints to benefit from the rule of reason would be to 'set sail on a sea of doubt'.

A fundamental widening of the scope of the rule of reason amounting to what is now known as the 'modern' rule of reason (Fasquelle, 1993) came with three cases, *Standard Oil* (1911), *American Tobacco* (1911) and *Board of Trade of Chicago* (1918)[97]. In *Standard Oil* the Surpreme Court explicitly narrowed down the scope of application of Section 1 to 'every undue or unreasonable restraint of trade'. In *American Tobacco*, the Surpreme Court clarified that the term 'restraint of trade' only embraces acts or contracts which operate 'to the prejudice of the public interest by unduly restricting competition' or which 'either because of the inherent nature or effect or because of the evident purpose of the acts' injuriously restrain trade. Finally, in *Board of Trade of Chicago*, it stated 'The true test of legality is whether the restraint imposed is such as merely regulates and perhaps thereby promotes competition or whether it is such as may suppress or even destroy competition. To determine that question, the court must ordinarily consider the facts peculiar to the business to which the restraint is applied; its condition before and after the restraint was imposed; the nature of its restraint and its effect actual or probable. The history of the restraint, the evil believed to exist, the purpose and end sought to be attained, are all relevant factors. This is not because of a good intention will save an otherwise objectionable regulation of the reverse; but because knowledge of intend may help the court to interpret the facts and to predict consequences'. The main thrust of these judgments - the emphasis on competition and the proposed test of legality - are generally regarded as the origin of the current American rule of reason, one that entails the determination of the 'competition balance' of an agreement to see whether it is pro- or anticompetitive.

The actual implementation of the American rule of reason has changed remarkably over the years. In order to avoid confusion, it must be noted that some categories of agreements, while subject to the rule of reason, have been viewed as 'per se unreasonable' (Van Gerven e.a., 1997; Fasquelle, 1993). In this respect, most changes in attitude have occurred in the domain of vertical restraints. The benchmark case is the *Sylvania* case of 1977[98]. In that case, the Supreme Court rejected the per se condemnation of purely vertical restraints in favour of a full rule of reason standard. In particular, it adopted the view that vertical restraints can have the 'redeeming virtues' of helping manufacturers increase interbrand competition by solving free-rider prob-

[97] *Standard Oil Co of New Jersey v. United States* 221 U.S. 1 (1911), *United States v American Tobacco* 221 U.S. 106 (1911), *Board of Trade of Chicago v. United States* 246 U.S. 231 (1918).
[98] *Continental TV Inc. v GTE Sylvania* 433 U.S. 36 (1977).

lems (Salop, 1993). The ruling reversed the 'per se unreasonable' approach adopted in the 1967 *Schwinn* case, which in turn had reversed the previous full rule of reason standard[99].

Since the *Sylvania* decision, economic efficiency has become the main substantive standard of US antitrust law (in line with the ideas of the Chicago School): agreements or business practices which restrict competition are those which reduce economic efficiency. In this regard, the emphasis has come to lie on the effects on price and output levels (allocative efficiency). Even when companies achieve considerable cost savings (productive efficiency), the question remains whether some of these savings are passed on to consumers through price reductions (Hirsch, 1979; Baker, 1999).

[99] *United States v Arnold Schwinn and Co* 388 U.S. 365 (1967).

3.9 References

Baker, J. (1999) 'Developments in antitrust economics' Journal of Economic Perspectives 13(1): 181-194.

Barendrecht, J. and Van Peursem, G. (1997) 'Distributieovereenkomsten' Recht en Praktijk, Kluwer, Deventer, Netherlands.

van den Bergh, R. (1997) 'Economische analyse van het mededingingsrecht' Rechtseconomische Verkenningen, Gouda Quint - Kluwer, Deventer, Netherlands.

Bishop, S. and Walker, M. (1999) 'The economics of EC competition law: concepts, applications and measurement', Sweet & Maxwell, London, UK.

van Damme, E. (1998) 'Verticale mededingingsafspraken' Sociaal Economische Wetgeving 1, 1-17.

Deacon, D. (1995) 'Vertical restraints under EU competition law: new directions' in Hawk (ed.): Annual Proceedings of the Fordham Corporate Law Institute 1995 - International Antitrust Law & Policy, Fordham Corporate Law Institute, New York, USA.

Ehlermann, C.-D. (2000) 'The modernisation of EC antitrust policy: a legal and cultural revolution' Common Market Law Review 37: 537-590.

Fasquelle, D. (1993) 'Droit américain et droit communautaire des ententes: étude de la règle de raison' Joly Editions, Paris, France.

Faull, J. and Nikpay, A. (1999) 'Article 81' in: Faull and Nikpay (eds.) 'The EC law of competition' Oxford University Press, Oxford, UK.

Freeman, P. (1996) 'New developments in vertical restraints', mimeo, Simmons and Simmons, London, UK.

van Gerven, W., Gyselen, L., Maresceau M. and Stuyck J. (1997) 'Kartelrecht Europese Gemeenschap' Tjeenk Willink - Kluwer, Deventer, Netherlands.

Gonzalez-Diaz, F.E. (1995) 'Some reflections on the notion of ancillary restraints under EC competition law' in Hawk (ed.): Annual Proceedings of the Fordham Corporate Law Institute 1995 - International Antitrust Law & Policy, Fordham Corporate Law Institute, New York, USA.

Goyder, D. (1993) 'EC competition law' Clarendon Press, Oxford, United Kingdom.

Goyder, D. (1996) 'EC distribution law' Wiley Chichester, United Kingdom.

Gyselen, L. (1984) 'Vertical restraints in the distribution process: strength and weakness of the free rider rationale under EEC competition law', Common Market Law Review 21: 647.

Griffiths, M. (2000) 'A glorification of *de minimis* - the Regulation on vertical agreements' European Competition Law Review 5: 241-247.

Hawk, B. (1989) 'La révolution antitrust américaine: une leçon pour la Communauté?' Revue Trimestrielle de Droit Européen 25(1): 5-45.

Hawk, B. (1995) 'System failure: vertical restraints and EC competition law' Common Market Law Review 4: 973-89.

Heinemann, A. (1999) commenting on Thomas Ackermann 'Art.85 Abs.1 EGV und die rule of reason' Common Market Law Review 37: 859-861.

Hildebrand, D. (1998) 'The role of economic analysis in the EC competition rules' Kluwer, The Hague, Netherlands.

Hirsch, W. (1979) 'Law and economics: an introductory analysis' Academic Press Inc., New York/London.

Kapteyn, P. and VerLoren van Themaat P. (1995) 'Inleiding tot het recht van de Europese Gemeenschappen' Kluwer, Deventer, the Netherlands.

Kmiecik, A. (2000) 'Comments on Regulation 2790/1999 on vertical restraints' Paper presented at the IBC Conference 13-14 November 2000, Brussels.

KORAH, V. (1981) 'The rise and fall of provisional validity - the need for a rule of reason in EEC antitrust' Northwestern Journal of International Law and Business 3: 320-357.

KORAH, V. (1998) 'The future of vertical agreements under EC competition law' European Competition Law Review 19, 506-512.

LEVER, J. AND NEUBAUER, S. (2000) 'Vertical restraints, their motivation and justification' European Competition Law Review 21(1), 7-23.

MARENCO, G. (1999) 'La notion de restriction de concurrence dans le cadre de l'interdiction des ententes' Mélanges en hommage de Michel Waelbroeck, Bruylant, Brussels.

NAZERALI, J. AND COWAN, D. (1999) 'Reforming EU distribution rules - has the Commission found vertical reality?' European Competition Law Review 20(3), 159-168.

PAULIS, E. AND PEEPERKORN, L. (1997) 'Green Paper on vertical restraints; new interpretation of Art. 85(1)?' Paper presented at the IBC Conference, Brussels, November 1997.

PEEPERKORN, L. (1999) 'The Commission's radical overhaul of EC competition policy towards vertical restraints' in: Faull and Nikpay (eds.) 'The EC law of competition' Oxford University Press, Oxford, UK.

RITTER, L., BRAUN, W. AND RAWLINSON, F. (1993) 'EEC competition law; a practioner's guide' Kluwer, Deventer-Boston.

SMITH, A. (1776) 'An inquiry into the nature and causes of the wealth of nations' Modern Library Edition, 1937, New York, USA.

SOUTY, F. (1996) 'La politique de la concurrence en Allemagne fédérale' Presses Universitaires de France, Paris, France.

SALOP, S. (1993) 'Exclusionary vertical restraints law: has economics mattered ?' American Economic Review 83 (2): 168-172.

WAELBROECK, M. (1987) 'Antitrust analysis under Article 85(1) and 85(3)' in Hawk (ed.): Annual Proceedings of the Fordham Corporate Law Institute 1987 - International Antitrust Law & Policy, Fordham Corporate Law Institute, New York, USA.

WESSELING, R. (1999) 'The Commission White Paper on modernisation of E.C. antitrust Law: unspoken consequences and incomplete treatment of alternative options' European Competition Law Review 20(8), 420-433.

WOLF, D. (1999) 'Perspektiven des Europäischen Kartellrechts' Positionspapier vorgelegt anläßlich des Diskussionsforums des Frankfurter Instituts - Stiftung Marktwirtschaft und Politik am 8. Juli 1999 in Frankfurt.

WILS, W. (2000) 'The modernisation of the enforcement of Articles 81 and 82 EC: a legal and economic analysis of the Commission's proposal for a new Council Regulation replacing Regulation No. 17' Paper presented at the Fordham Corporate Law Institute 27th Annual Conference on International Antitrust Law & Policy, 19-20 October 2000, New York, USA.

WOODS, D. AND FILIPPONI, M. (1999) 'Vertical agreements' in: Faull and Nikpay (eds.) 'The EC law of competition' Oxford University Press, Oxford, UK.

Chapter 4

Resale price maintenance in a spatial market with fixed transportation costs

4.1 Introduction

A contentious topic, in antitrust and in economics, continues to be the use of (minimum) resale price maintenance. Under this practice, a manufacturer requires retailers not to sell its products below a certain minimum price. Equally, a manufacturer's unilateral policy not to deal with 'discounters' can be considered a kind of resale price maintenance.

Several explanations for the use of resale price maintenance (RPM) have been given in the economic literature. As has been discussed in Chapter 2, there are, broadly speaking, two opposing views on the impact of resale price maintenance on competition which can be identified. The malign view regards such an arrangement as anti-competitive, serving to reduce intrabrand competition (competition among distributors of the same brand) or even interbrand competition (competition between brands). The alternative view is that minimum resale price maintenance can resolve certain contracting or principal-agent problems, thereby enhancing efficiency in distribution. For example, RPM can be used to prevent discount retailers from free-riding on product-specific presale services provided for by traditional dealers.

One efficiency rationale that has received little recent attention is the so-called *outlets hypothesis,* articulated by Yamey (1954) and elaborated upon by Gould and

Preston (1965). The outlets hypothesis assumes that final demand for the manufacturer's product is a function both of the retail price and the number of retail outlets: the price-demand schedule for a product shifts outward if the number of retailers carrying the product increases. One of the informal arguments for this positive relationship is that the inconvenience of shopping (e.g. travelling) is reduced when retail density is higher. Gould and Preston then argue that price floors, by raising the retail margin above the competitive level, lead to an increase in the number of retail outlets in a free entry market equilibrium and, because of the positive impact of this increase on final demand, to higher profits for the manufacturer.

It is not entirely clear why the outlets hypothesis seems to have dissappeared from the foreground of antitrust economics. According to Reagan (1986), this oversight is explained partly by the lack of a formal model supporting the claim that minimum resale prices can be profitably used by a manufacturer to establish its preferred retail outlet density. It is even the case that two articles which have evaluated Gould and Preston's argument in a spatial context, Mathewson and Winter (1983) and Bittlingmayer (1983), provide theoretical support only for the use of *price ceilings,* not for price floors! Their results are obtained using a model of spatial retail competition in which the number of retailers and their locations are fixed in the short run but variable in the long run. More specifically, their models consist of three stages: in the first stage, a monopolistic producer decides on the sales conditions vis-à-vis the retailers (wholesale price, possible minimum or maximum prices and/or franchise fees). In the next stage, retailers decide whether or not to establish a retail outlet in a (circular) city, given the sales conditions of the producer. Retailers will enter the market if the prospective profits from sales are sufficient to cover the fixed costs associated with setting up the retail outlet[1]. In the last stage, the established retailers compete with each other in prices. The (fully informed) consumers are assumed to have a downward sloping demand curve and to buy at the retailer whose 'effective price' (the sum of the retail price and the unit transportation cost, which depends on the distance to the retail outlet) is lowest.

In determining the optimal number of retail outlets, the manufacturer has two

[1] A tempting interpretation of the fixed costs is that they represent the opportunity costs of shelf space (profits foregone by having the product concerned in the shelves and not some other product). However, this interpretation implies a multiproduct setting, a setting to which the model developed in this chapter does not seem very applicable. In particular, the symmetry assumption, needed for a tractable analysis, will become very difficult to justify.

opposite effects to take into account. Firstly, the higher the number of retail outlets entering the market in equilibrium, the higher the sum of fixed costs involved (to be covered through the retail mark-up). On the other hand, an increase in the number of retailers may go with a decrease in the 'effective price' faced by consumers as the average travelling distance for consumers decreases. For a given wholesale price, a drop in the 'effective price' benefits the producer as the total quantity of goods sold increases. Therefore, the reduction in transportation costs incurred by the consumers may allow the producer to capture a larger part of the consumer surplus. In the constellations of Mathewson and Winter (1983) and Bittlingmayer (1983), the first effect - as evaluated at the equilibrium without vertical restraints - appears to always dominate the second effect. The essence of this result is that in the absence of vertical restraints, there is a strong 'business stealing effect' (Tirole, 1988): retailers, when deciding to enter the market, do not take into account the negative effect of entry on the profits of the other retailers. From the viewpoint of the vertical structure, this leads to a certain bias towards excess entry. It also renders the result that price ceilings, and not price floors, are called for if the producer wants to maximize profits by influencing the number of retail outlets in the long run[2]. In a different constellation, that of Vickrey(1964)/Salop (1979)/Dixit (1983) with inelastic demands, it is shown that the producer is able to achieve the optimal outlet density by using the wholesale price only. These findings have led Reagan (1986) to the conclusion that there is 'no substantive role for minimum resale prices in the long-run'[3].

In this chapter, I will verify whether the above 'negative' results for the outlets hypothesis are due to the particular transportation cost assumptions of the underlying models. In line with the majority of spatial models of retail competition, these models have assumed that transportation costs depend linearly on the quantity of products actually bought at the retail outlet. By contrast, I will assume that customers only incur a fixed cost when visiting a retail outlet, i.e. a cost irrespective of whether they buy several products or nothing at all. This assumption, first put forward in spatial models by Holton (1957) and supported by Stahl (1982, 1996), seems quite justified when the size or quantity of the goods transported is relatively small (in terms of

[2]Note that the use of a price ceiling is also beneficial in view of the double marginalisation problem. As equally effective instruments exclusive territories or two-part tariffs can be used.

[3]As retailers choose whether or not to enter the market before price competition takes place, the entry decision can be interpreted as a decision for the long run. Reagan does provide an interesting short term model (without the possibility of entry, taking the locations of the retailers as given), with which she derives a result supporting the use of price floors.

travel expenses it does not seem to matter a lot whether a customer buys, say, either one or two bottles of coke, or either one or two packets of aspirine in a drug store) or transportation costs are looked upon as opportunity costs associated with the time spent on shopping and not on other activities (after all, it is not obvious why buying several products at the same time would generally lead to more inconvenience than buying just a single product).

An interesting feature of the fact that transportation costs are not dependent on the quantity bought is that it changes the analysis of where a consumer will buy. The larger the quantity consumers want to buy, the more they will get interested in going to a remote shop with a low price, rather than going to the shop closeby with high prices. In a way, the choice of where to buy also depends on the choice how much to buy. The fixed transportation cost structure thus appears to bring about more competitive retail conditions than the linearly dependent transportation cost structures. In principle, this may reduce the extent of the excess entry bias of spatial models and, therefore, lead to price floors being more attractive as a means to foster entry by retailers.

However, despite this feature it is found that also with the fixed transport cost specification, the producer does not find it profitable to impose price floors. Again, the better capture of consumer surplus appears not to weigh up against the increase in fixed costs involved with the larger retail network. As in Mathewson and Winter (1983) and Bittlingmayer (1983), the business stealing effect (retailers enter without taking into account the effect on the other retailers in terms of reduced profits) is determinative: an integrated firm internalizes this externality by reducing the number of retail outlets. The co-ordination problems that exist within the vertical structure (the double marginalisation problem and the co-ordination problem on market coverage) can either be tackled using a different pricing instrument (price ceilings, two-part tariffs) or are not strong enough to offset the business stealing effect.

The next section, Section 4.2, presents the assumptions of the model and characterises the retail market equilibria. Section 4.3 derives the equilibrium for the case where the producer does not impose vertical restraints and for the case it does. In Section 4.4, the welfare implications of the model are considered. Concluding comments and a short discussion are offered in Section 4.5.

4.2 The basic model

The role of resale price maintenance will be studied in a spatial competition framework with free entry by retailers. I will use a constellation similar to Mathewson and Winter (1983), Bittlingmayer (1983) and Dixit (1983) but introduce one important distinctive feature regarding the nature of the transporation cost.

The model consists of three stages:

- stage I: the manufacturer decides on the (uniform) sales conditions vis-à-vis the retailers, i.e. on a wholesale price p_w, a possible franchise fee A and on whether or not to use price restrictions (price floors, price ceilings or full resale price maintenance);
- stage II: given the manufacturer's choice in stage I, retailers decide simultaneously whether or not to enter the market;
- stage III: given the retailers' locations, retailers compete in prices.

The following assumptions are made, some of which are standard in the spatial competition literature.

(i) Consumers are uniformly distributed with density v along a circle with circumference 1. Hence, spatial retail competition is in one dimension only;

(ii) Consumers have a common travel cost t per unit distance.

(iii) The manufacturer of the product enjoys some market power in the sense that it faces a downward sloping consumer demand for its product. It will be convenient to think of the manufacturer as a monopolist. The manufacturer incurs a constant unit cost of production, c. There are no costs involved in the transportation of the goods to the retail outlets.

(iv) Each retailer incurs a fixed cost F associated with setting up the retail outlet. There are no variable costs in retailing (apart from purchasing the products at the wholesale price p_w).

(v) There is free entry into the retail market, implying that retailers will continue to enter the market until profits are driven down to zero. The number of retailers entering is, for the moment, supposed to be large enough to be treated as a continuous variable.

(vi) Retailers locate symmetrically along the retail market[4].

(vii) Each consumer buys at the retail outlet where its utility ends up highest[5], provided that the utility level is positive. Utility increases in the quantity of products obtained but decreases in the total expenses made (being the sum of buying expenses and travel expenses). To keep things as simple as possible, it will be assumed that consumers have a common utility function that is quasi-linear:

$$U(q,x) = u(q) - p \cdot q - t \cdot x,$$

Here, q denotes the quantity of goods obtained, $u(q)$ the utility associated with consuming this quantity, x the distance from the consumer's location to the retail outlet where the products are bought, and p the retail price at the retail outlet. As is well known, quasi-linear utility functions have the characteristic that income effects are absent, which allows us to write demand of a consumer (who decides to visit a retail outlet) as a function of the retail price only[6]. In line with Mathewson and Winter (1983), we take an exponential demand function: $q(p) = e^{-p}$, which is in fact equivalent to taking the specification $u(q) = q - q \cdot \ln q$. Hence, if a consumer located at a distance x from a retail outlet buys from this outlet, he will derive a utility level equal to

$$\begin{aligned} U(q(p),x) &= e^{-p} - e^{-p} \cdot (-p) - p \cdot e^{-p} - t \cdot x \\ &= e^{-p} - t \cdot x \end{aligned}$$

Observe that this expression is identical to the difference between the (net) consumer's surplus of consuming $q(p)$ at a price p and the transportation cost[7].

[4]Obviously a model in which firms choose whether or not to enter and also the locations at which they will enter would be more appropriate. However, in order to keep computations tractable, it will prove necessary to make the assumption. Furthermore, the specific objective of the model is to study the extent of entry rather than the particular choice of locations (cf. Tirole, 1988).

[5]In the setting of Mathewson and Winter (1983) and of most other works, it is assumed that the transportation cost t is per unit distance per unit quantity purchased. This implies that a consumer at a distance x from a retail outlet faces a (constant) effective retail price equaling $\bar{p} = p + t \cdot x$. Maximizing utility then coincides with going to the retail oulet whose effective retail price (or 'delivery price') is lowest.

[6]The absence of income effects only applies as long demand for a product is not constrained by a (low) income level. Hence, it is assumed that income is large or, equivalently, expenditures made on the product under consideration are small relative to the income level.

[7]Allowing for different positive demand elasticities, e.g. by taking the specification $q(p) = e^{-ap}$, $q(p) = a - p$ or $q(p) = p^{-\epsilon}$ $(a, \epsilon > 0)$, yields identical results.

On the basis of the above elements, it is now possible to derive the perceived demand curves for a single retailer, conditional on the prices set by the other retailers. Three possible price segments of a retailer's perceived demand curve can be distinguished: following Salop (1979), I will call them the 'monopoly' segment, the 'competitive' segment and the 'supercompetitive' segment. The monopoly segment consists of prices at which the retailer's entire market area consists of consumers who are only interested in his shop as a visit to any other shop would yield negative utility (because of the high prices charged by the other shops or because the other shops are simply located too far away). The competitive price segment of the demand schedule consists of those prices at which also consumers who are, in principle, willing to go to other shops, decide to come to his retail outlet. For these consumers the retailer will have to compete directly with the other retailers. This is equivalent to saying that the potential market areas (the market areas that could be served if there were no competitors) of the retailers are overlapping. Because of the interfering presence of other retailers, demand on the competitive segment will start reacting less elastically on price decreases, rendering a 'kink' in the demand curve at the price at which the market areas of the retailers start to touch each other. The supercompetitive price segment, finally, is related to prices which are that low that the immediate competitors of the retailer are left with no customers at all (this segment will be ignored for the moment). In Figure 4.1, the two typical segments of a retailer's demand curve are depicted.

Let us derive the demand schedules for the two price segments in turn. Suppose that retail outlets are located symmetrically at a distance R from each other. Consider a firm i whose retail price is p_i and let the price of retailer i's neighbouring rivals be p_{-i}. As mentioned, on the *competitive segment* of the demand curve it will be the case that the retailer competes with the neighbouring retailers for consumers who will choose that retail outlet where their utility ends up highest. The market area of retailer i is then given by the location $\bar{x}$ of the so-called 'indifferent consumer', who is indifferent between buying at i and buying at its neighbour:

$$U(q(p_i), \bar{x}) = U(q(p_{-i}), R - \bar{x})$$

or

$$e^{-p_i} - t \cdot \bar{x} = e^{-p_{-i}} - t(R - \bar{x})$$

leading to

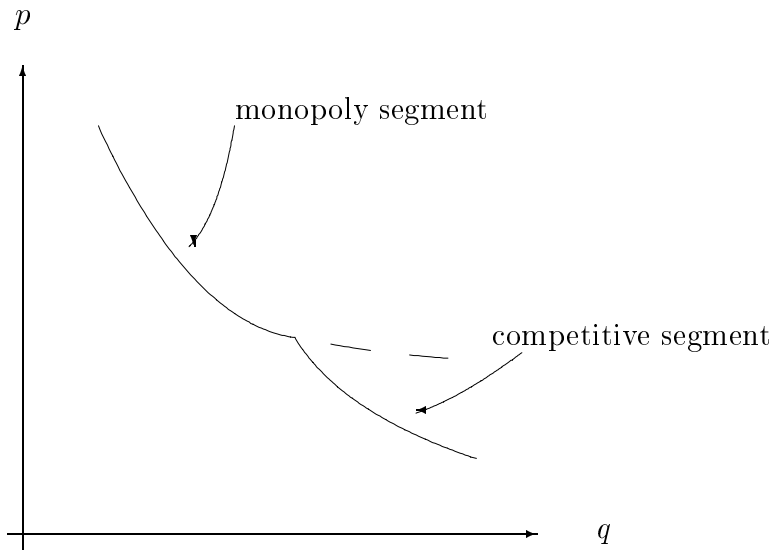

Figure 4.1: The typical perceived demand curve for a retailer.

$$\bar{x} = \frac{1}{2t} \cdot (e^{-p_i} - e^{-p_{-i}} + tR)$$

As the indifferent consumer derives non-negative utility from visiting a retail outlet ($U(q(p_i), \bar{x}) \geq 0$), consumers located nearer to the outlet will do so as well and, hence, market demand faced by retailer i on the competitive price segment is given by

$$\begin{aligned} q^c(p_i, p_{-i,}R) &= 2 \cdot \int_0^{\bar{x}} v \cdot e^{-p_i} dx \\ &= 2v \cdot \bar{x} \cdot e^{-p_i} \\ &= \frac{v}{t} \cdot (e^{-p_i} - e^{-p_{-i}} + tR) \cdot e^{-p_i} \end{aligned}$$

On the *monopoly segment* of the demand schedule, retailer i serves a market area which entirely consists of consumers who derive negative utility from going to other shops. Then, the location of the consumer who is just willing to purchase goods from i, $x_{\max}$, defines the boundary of the market area served by retailer i. From $U(q(p_i), x_{\max}) = 0$, it follows that $x_{\max} = (1/t) \cdot e^{-p_i}$ so that in the case of separated market areas demand for retailer i is equal to

$$\begin{aligned} q^m(p_i, R) &= 2 \cdot \int_0^{x_{\max}} v \cdot e^{-p_i} dx \\ &= 2v \cdot x_{\max} \cdot e^{-p_i} \\ &= \frac{2v}{t} \cdot e^{-2p_i} \end{aligned}$$

In the next section we will derive retail market equilibria given that the manufacturer sells its products at a wholesale price p_w and does not impose any other sales conditions on the retailers.

4.2.1 Conditions characterising the possible types of equilibria

A symmetric free entry retail market equilibrium is characterised by the requirement that (i) prices and locations are symmetric, (ii) given the prices of the competitors, the best a retailer can do is to stick to its price and (iii) each retailer earns exactly a zero profit.

Three possible types of equilibrium configurations can be distinguished (see also Figure 4.2). The first type is the 'competitive equilibrium', i.e. the equilibrium configuration in which the potential market areas are overlapping. The second type is the '(free entry) monopoly equilibrium', where it is the case that retailers are not directly competing with each other for customers (the potential market areas are not overlapping). The third type are the 'kinked equilibria', where the prices are such that the market areas (the actual as well as the potential market areas) just touch. These equilibria involve corner solutions in the profit maximization problem of the retailers. It will turn out that for low levels of the wholesale price, the resulting retail equilibria are the competitive equilibria, for intermediate levels the resulting equilibria are the kinked equilibria and that for a high wholesale price a free entry monopoly equilibrium will be the result.

We will start with a characterisation of the competitive equilibria. For expositional reasons, part of the maths underlying the respective propositions is outlined in the text. The details of the proofs are relegated to the appendix.

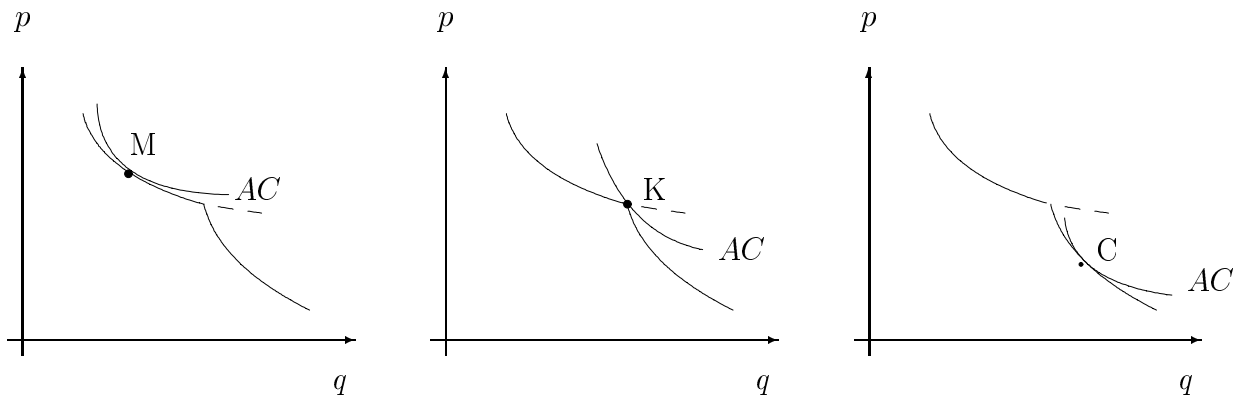

Figure 4.2: The three possible types of retail equilibria (curve AC represents the average cost curve).

The competitive equilibria

In a competitive equilibrium, the potential market areas of the retailers turn out to overlap. Suppose that the manufacturer has set its sales conditions in stage I and that n retailers have entered in stage II. In the last stage, stage III, these retailers compete in prices, given each other's locations and given the producer's wholesale price. In a Nash price equilibrium, each retailer maximizes its profits given the prices of the other retailers:

$$\pi^c(p_i, p_{-i,}R) = (p_i - p_w) \cdot q^c(p_i, p_{-i,}R) - F$$

The first order condition of profit maximization is

$$\frac{\partial}{\partial p_i}\pi^c(p_i, p_{-i,}R) = 0$$

Assuming symmetry in prices, $p_i = p_{-i} = p$, the first order condition boils down to:

$$(p - p_w) \cdot (e^{-p} + tR) - tR = 0 \tag{4.1}$$

which determines equilibrium price p as an (implicit) function of p_w, the wholesale price, and R, the distance between the retailers. The corresponding profits are then

$$\pi^c(p, R) = R \cdot v \cdot e^{-p} \cdot (p - p_w) - F$$

In stage II, the stage preceding the above price competition stage, retailers choose whether or not to enter the market. How large the distance between shops, R, will turn out to be is determined by the number n of retailers deciding to enter: $R = \frac{L}{n}$. As we have assumed free entry, the number of retailers will be such that the above profits are driven down to zero:

$$\pi(p, R) = R \cdot v \cdot e^{-p} \cdot (p - p_w) - F = 0 \tag{4.2}$$

Equations (4.1) and (4.2) determine, implicitly, the unique combination of price and retail area $(p(p_w), R(p_w))$ in the free entry competitive equilibrium[8], for each wholesale price p_w:

$$\begin{cases} v \cdot e^{-2p} \cdot (p - p_w)^2 + tF \cdot (p - p_w) = tF \\ R = \frac{F \cdot e^p}{(p - p_w) \cdot v} \end{cases} \tag{4.3}$$

It is fairly straightforward to observe, by implicitly differentiating the above expressions, that as the wholesale price p_w increases, both the retail price and market area of each retailer in the competitive equilibrium increase. Hence, if the wholesale price increases, outlet density in the competitive case decreases.

That only implicit results can be obtained is slightly unfortunate, but appears almost inevitable in models of this kind. In any case, it will prove useful to identify, by eliminating wholesale price p_w from equations (4.3), that the set of possible combinations of retail price p and retail area R is given by the following relation (f)

$$p = \ln(\frac{vR}{F} - \frac{1}{tR}) =: f(R)$$

By differentiating this expression we observe that in a competitive equilibrium, price and retail area are positively related $(f'(R) > 0)$ and that f is concave in R $(f''(R) < 0)$. The intuition is that as the wholesale price increases, firms will, given the retail price, no longer be able to recover their fixed costs and, hence, the outlet density that is sustainable in equilibrium will decrease. Both the higher wholesale price and the reduction in the degree of price competition (due to larger distances between outlets) have the effect of raising the equilibrium retail price. Obviously, this cannot go on indefinitely: if the values for retail price p and retail area R implied by expression (4.3) become too high, the consumer located furthest away from the retail outlet, the indifferent consumer, will derive negative utility from visiting the outlet and hence

[8]Locally and globally, the second order condition is satisfied: see appendix.

not come to the outlet at all. Therefore, for relation f to be a valid representation of the competitive equilibria, it must be that:

$$U(q(p), \bar{x}(p)) \geq 0$$

Note that, with symmetry, $\bar{x} = \frac{1}{2}R$, so that the requirement translates into

$$p \leq \ln \frac{2}{tR} =: g(R)$$

Dixit (1983) calls this condition the requirement of 'compatibility with demand'. It simply tells us that for a given price, the market area that can be served is limited. Accordingly, the interpretation of the inverse function g^{-1} is that it equals, for a given price, the potential market area of a retailer (being two times the distance of the consumer who derives exactly zero utility from going to the shop, i.e. $2 \cdot x_{\max}(p)$). Function g is downward sloping: the larger the retail price, the smaller the potential market area.

Figure 4.3 provides a plot of the curves f and g in order to exhibit the cases in which combinations (p, R) on f are indeed possible competitive equilibria. The requirement of compatibility with demand implies that the only relevant competitive equilibrium combinations (p, R) are beneath curve g.

Observe that the price-area relation f is parameterized by v, t and F: the price-area curve shifts outward when F increases, v decreases or t decreases. If fixed costs F increase, then given the retail price, the quantity sold (the retail area) will have to increase for the shops still to break even. With smaller consumer density v, the retail area will also have to increase as demand from the original retail area does no longer suffice. If the cost of transportation t decreases, then, given the locations of the shops, retail competition will become more intense, causing a loss to the retailers. The zero-profit condition then implies enlarged retail areas. The requirement of 'compatibility with demand' g is only parameterized by the transportation cost t: if t increases, the potential market area decreases (for given prices).

In each competitive equilibrium the market is fully covered and the potential market areas are overlapping ($f^{-1}(p) < g^{-1}(p)$). At the point at which the curves f and g intersect, the potential market areas just touch each other. At this point, which we will call S, it holds that $f(p) = g(p)$ and we have

$$R^S = \sqrt{\frac{3F}{vt}}$$

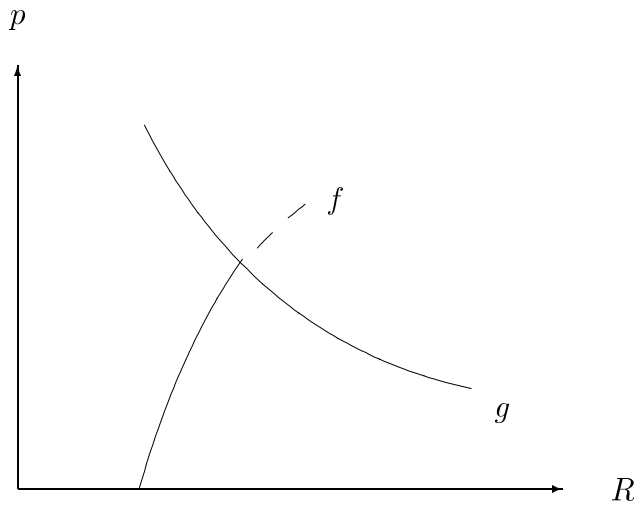

Figure 4.3: Price-area relation f and the requirement of compatibility with demand g.

$$p^S = \ln 2\sqrt{\frac{v}{3Ft}}$$

As mentioned above, it is straightforward to observe that as the wholesale price increases, both the retail price and the market area in the competitive equilibrium increase. Therefore, the maximum wholesale price that still leads to a competitive equilibrium is the wholesale price that supports equilibrium[9] S. Inserting the values p^S and R^S into zero profit condition (4.2), gives that

$$p_w^S = \ln 2\sqrt{\frac{v}{3Ft}} - \frac{2}{3}$$

The findings of this section are summarized in the following lemma.

Lemma 4.1 *For each wholesale price $p_w < p_w^S$, there is a unique competitive equilibrium (p, R) that satisfies zero profit condition (4.2), $p = f(R)$ and $p < g(R)$. In this equilibrium, the number of firms entering is $\frac{L}{R}$.*

Proof: see appendix.

[9]Strictly speaking, S itself is no longer a competitive equilibrium, but a kinked equilibrium (the equilibrium potential market areas are adjacent; see section 2.1.3). Therefore, the obtained p_w^S is in fact a supremum value for the wholesale prices leading to competitive equilibria.

The free entry monopoly equilibrium

In a free entry monopoly equilibrium, retailers turn out to be not directly competing with each other. Profits for a local monopolist if it sets price p are

$$\pi_i^m(p) = q^m(p) \cdot (p - p_w) - F$$

The first order condition of profit maximization, $\frac{d}{dp}\pi_i^m(p) = 0$, gives[10]

$$p^m(p_w) = p_w + \frac{1}{2} \tag{4.4}$$

The corresponding profits are

$$\pi_i^m(p^m(p_w)) = \frac{v}{t} \cdot e^{-2p_w - 1} - F$$

which are a decreasing function of p_w. The only free entry monopoly equilibrium which is compatible with the zero profit requirement is the equilibrium corresponding to

$$p_w = \ln(\sqrt{\frac{v}{Ft}}) - \frac{1}{2} =: p_w^M$$

Wholesale price p_w^M can be seen as the maximal wholesale price at which a retailer who is able to act as a local monopolist can still break even. The retail price prevailing in the (unique) monopoly equilibrium M is

$$p^M = \ln(\sqrt{\frac{v}{Ft}})$$

with corresponding retail area

$$R^M = 2x_{\max}(p^M) = 2\sqrt{\frac{F}{vt}}$$

These results are briefly summarized in the following lemma.

Lemma 4.2 *The unique free entry monopoly equilibrium M is given by $p^M = \ln(\sqrt{v/Ft})$ and $R^M = 2\sqrt{F/vt}$ and occurs for $p_w = p_w^M$.*

Proof: see appendix.

[10] The second order condition is satisfied: see the appendix.

The kinked equilibria

The third type of equilibria is the 'kinked equilibria'. In a kinked equilibrium, the prices and market areas are such that, starting from the equilibrium, any increase in the price brings demand on the monopoly segment (i.e. the potential market areas become separated), and any price decrease brings demand on the competitive segment (i.e. potential market areas start to overlap). Because of this feature, the first order condition of profit maximization must hold as an inequality. This requirement and the requirement that each retailer makes zero profit characterise the kinked equilibria. As in a kinked equilibrium the potential market areas just touch each other, it is implied that $R^k = 2 \cdot x_{\max}(p^k)$, which is equivalent to $p^k = g(R^k)$. Combining this with the zero profit condition, $\pi^k(p, R) = R \cdot v \cdot e^{-p}(p - p_w) - F = 0$, we obtain:

$$\frac{1}{2} vtR^2 (\ln \frac{2}{tR} - p_w) - F = 0 \tag{4.5}$$

Expression (4.5) defines implicitly kinked equilibrium market area R^k as a function of wholesale price p_w. Implicit differentiation of (4.5) reveals that when the wholesale price increases, the equilibrium market area increases as well, but the price level drops: in the region of the kinked equilibria, the induced increase in the market area goes along with a stronger exploitation of scale economies (lower prices), an effect that dominates the market power effect due to greater spatial differentiation. Furthermore, each firm in the kinked equilibrium would, in principle, prefer to decrease its price and charge the monopoly price, but is refrained from doing so because of the presence of other competitors; a price decrease enlarges the actual market area by less than the potential market area. In the limit, when the wholesale price goes towards p_w^M, the kinked equilibrium coincides with the free entry monopoly equilibrium. Correspondingly, market area R^k is smaller than area R^M. The next lemma summarizes the main results.

Lemma 4.3 *For each wholesale price $p_w \in [p_w^S, p_w^M)$, there is a unique kinked equilibrium (p, R) that satisfies $p = g(R)$ and zero profit condition (4.5). The number of firms entering is $\frac{L}{R}$.*

Proof: see appendix.

In Figure 4.4 the possible equilibria are depicted for three cases. For low wholesale prices, the resulting equilibria are the competitive equilibria (Lemma 4.1), for intermediate wholesale prices the resulting equilibria are kinked equilibria (Lemma 4.3) and for $p_w = p_w^M$ the free entry monopoly equilibrium will be the result (Lemma 4.2).

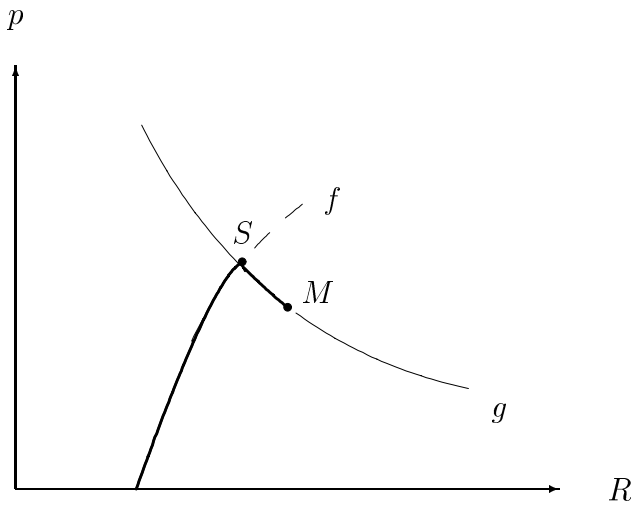

Figure 4.4: Possible retail equilibria; one for each wholesale price p_w.

4.2.2 A comparison with the case of per-unit transportation costs

In the introduction it was stated that two leading articles studying vertical integration in spatial models of retail competition (Mathewson and Winter (1983) and Bittlingmayer (1983)) assume that transportation costs depend linearly on the quantity of products actually bought at the shop. If t is the transportation cost per unit distance, per unit bought, this implies that a consumer at a distance x from a shop faces a (constant) effective retail price equaling $p + t \cdot x$. Maximizing utility then coincides with going to the shop whose effective retail price (or 'delivery price') is lowest. In the introduction, it was also asserted that the fixed transportation cost structure of the current model brings about more competitive retail conditions than the linearly dependent transportation cost structures. Let us quickly verify this assertion. For given retail locations (for given R), profit maximization in the per-unit transportation cost framework gives (cf. Mathewson and Winter (1983), p. 501):

$$p - p_w = \frac{2(1 - e^{-tR})}{(1 - e^{-tR})}$$

Note that in this framework, demand of a consumer located at a distance x equals quantity $ve^{-(p+t\cdot x)}$, which is positive for all distances x (in other words, $x_{\max}(p) = \infty$).

Hence, with per-unit transportation costs the market is always completely covered: there are only competitive equilibria, no kinked equilibria nor monopoly equilibria.

In the competitive equilibria of the fixed transportation cost set-up it holds that (expression (4.3))

$$p - p_w = \frac{tR}{(e^{-p} + tR)}$$

It is easily verified that since

$$\frac{tR}{(e^{-p} + tR)} < \frac{2(1 - e^{-tR})}{(1 - e^{-tR})}$$

the retail conditions are more competitive in the fixed transportation cost framework than in the framework with per-unit transportation costs, in the sense that for a given distance between retail outlets, the retail margins are lower. The same holds for parameter values for which the retail equilibria in the fixed cost set-up are either kinked or monopolistic. The next section will consider whether these more 'competitive' retail conditions reduce the extent of the excess entry bias of the spatial model and lead to price floors being a profitable means to foster entry by retailers.

4.3 The choice of the manufacturer

In stage I, preceding the entry decision of the retailers, the producer decides on the sales conditions vis-à-vis the retailers with the objective of maximizing its profits. In this section we explore the following instruments that the producer can use: the wholesale price, possible minimum or maximum prices and/or franchise fees. However, I shall first characterise the 'first-best' result, i.e. the result that maximizes profit for the vertically integrated structure. This will serve as a benchmark with which the performance of the respective instruments can be assessed.

4.3.1 Vertically integrated distribution

When the number of operating retailers is n, profits of the vertically integrated structure are

$$\begin{aligned} \Pi_{int}(p, R) &= vL \cdot q(p) \cdot (p - c) - nF \\ &= L[v \cdot e^{-p} \cdot (p - c) - F/R] \end{aligned} \tag{4.6}$$

provided that the market is covered, $p \leq g(R)$. Inspection of the profit function learns, that for each given price p, Π_{int} is strictly increasing in R, so that it is optimal for the vertically integrated firm to choose R as large as possible while assuring that the market remains covered. Hence, the optimum is to be found on the line $p = g(R)$. This result is not surprising: provided that consumers come to a shop, the quantity they buy is not influenced by the extent of their travel expenses. The best thing a vertically integrated structure can do is to minimize on the number of retail outlets (in order to save fixed costs) and to position the shops in such a way that the customers who are located furthest away from them derive exactly zero utility from purchasing.

The optimal value for the control variable R can be derived taking the first derivative of $\Pi_{int}(g(R), R)$ with respect to R and setting it to zero:

$$\frac{d\Pi_{int}(g(R), R)}{dR} = \frac{1}{2} tvL(\ln \frac{2}{tR} - c - 1) + \frac{F}{R^2} = 0 \tag{4.7}$$

This expression determines, albeit implicitly, the optimal market area for the integrated firm, R^I. The corresponding optimal price is $p^I = g(R^I)$.

4.3.2 Maximizing producer's profits in the absence of vertical restraints: the non-integrated optimum

If a producer cannot use vertical restraints or franchise fees, the only decision variable is the price at which it sells to the retailers, the wholesale price. Let $p(p_w)$ denote the price that arises in the retail market equilibrium for a given wholesale price p_w, $n(p_w)$ the number of entering retailers and $R(p_w)$ the corresponding market areas. The profits for the producer can be written as

$$\Pi_{prod}(p_w) = vL \cdot q(p(p_w)) \cdot (p_w - c)$$

(note that the fixed costs of the outlets are paid for by the retailers themselves). A property of the free entry retail market equilibrium is that profits at the downstream level are zero (otherwise more retailers would enter). This implies that, in each retail market equilibrium, the profits of the vertical structure seen as a whole (Π_{int}) fully

accrue to the producer:

$$\begin{aligned}\Pi_{prod}(p_w) &= vL \cdot q(p(p_w)) \cdot (p_w - c) \\ &= vL \cdot q(p(p_w)) \cdot (p(p_w) - c) - n(p_w) \cdot F \\ &\quad -n[vR \cdot q(p(p_w)) \cdot (p(p_w) - p_w) - F] \\ &= vL \cdot q(p(p_w)) \cdot (p(p_w) - c) - n(p_w) \cdot F \\ &= \Pi_{int}(p(p_w), R(p_w))\end{aligned}$$

as $vR \cdot q(p(p_w)) \cdot (p(p_w) - p_w) - F = 0$ in the free entry retail market equilibrium. This implies that the producer's profits are maximal when the downstream variables $(p(p_w), R(p_w))$ 'end up' being equal to (p^I, R^I), the optimal price-area combination for the vertically integrated structure. Hence, the question when the wholesale price is the only available instrument is: is there a wholesale price p_w such that the resulting pair $(p(p_w), R(p_w))$ equals (p^I, R^I)? If not, what is the best possible result? The answer to these questions is stated in the following propositions.

Proposition 4.1 *The non-integrated optimum is given by the free entry monopoly equilibrium for large fixed costs F and by a competitive equilibrium for small values of F.*

Proposition 4.2 *The retail price and outlet density of the non-integrated optimum are too high from the viewpoint of the manufacturer/the vertical structure (unless the fixed costs are that large that even a vertically integrated firm would only just be able to break even in its operations; then the non-integrated optimum and the integrated optimum coincide).*

Proofs. See appendix.

Recall from the previous section that three possible equilibria may arise. For low wholesale prices, the resulting equilibria are competitive equilibria, for intermediate wholesale prices the resulting equilibria are the kinked equilibria and for $p_w = p_w^M$ the free entry monopoly equilibrium will be the outcome.

Let us first compare the free entry monopoly equilibrium and the kinked equilibria. It follows immediately that the former dominates the latter from the viewpoint of the manufacturer: for all kinked equilibria consumer demand is lower than in the free entry monopoly equilibrium ($p^k > p^M$), whereas the wholesale margin for the manufacturer is lower as well ($p_w^k < p_w^M$). The kinked equilibria are therefore never optimal.

Let us now consider the free entry monopoly equilibrium with the integrated optimum (p^I, R^I). The free entry monopoly equilibrium resembles the integrated optimum in that the customers who are located furthest away from a shop are left with exactly zero utility (it holds that $p = g(R)$). In terms of the appropriation of consumer surplus, this speaks in favour of the free entry monopoly equilibrium. Nonetheless, because of the double marginalisation problem, each monopoly equilibrium goes with higher prices and, hence, smaller retail areas than is optimal for the vertical structure as a whole. The only exception is when $p_w^M = c$, in which case there is no double marginalisation problem and the free entry monopoly equilibrium attains the first best. As p_w^M is, by definition, the highest possible wholesale price at which a local monopoly is still able to break even, the case $p_w^M = c$ occurs when the fixed costs are that large that even a vertically integrated firm would only just be able to break even in its operations (we will call this level F_{viable}). The conclusion is that whenever a free entry monopoly equilibrium is optimal for the manufacturer it will involve a retail price and, correspondingly, an outlet density that is too high (unless $F = F_{viable}$).

What remains is to compare the overall profits in free entry monopoly equilibrium M with those of the best possible competitive equilibrium, which we call C. Equilibrium C is represented in Figure 4.5 by the point where the iso-profit curve of the vertical structure (not drawn) touches the segment of possible competitive equilibria.

As with the kinked equilibria, we can readily say that all competitive equilibria that lead to a retail price above the monopoly equilibrium price p^M are less profitable than the free entry monopoly equilibrium: consumer demand is lower and so are the wholesale margins. One is left to consider competitive equilibria associated with low retail prices. Can they be more profitable than the free entry monopoly equilibrium? Recall again the definition of wholesale price associated with the monopoly equilibirum, p_w^M. This wholesale price is, by definition, the highest possible wholesale price at which a local monopoly is still able to break even. It follows that when the fixed costs are (very) small, having the retailers obtain a zero profit in a monopoly situation requires (very) high wholesale prices p_w. In fact, when $F \approx 0$, both p^M and p_w^M go towards infinity, leaving zero profit for the manufacturer. Hence, even though the free entry monopoly equilibrium resembles the industry optimum in that some customers are left with zero utility, it simply requires too high wholesale and retail prices when the costs of setting up a shop are, in fact, low. When the fixed costs of setting up a retail outlet are low, it will be a competitive equilibrium that is optimal.

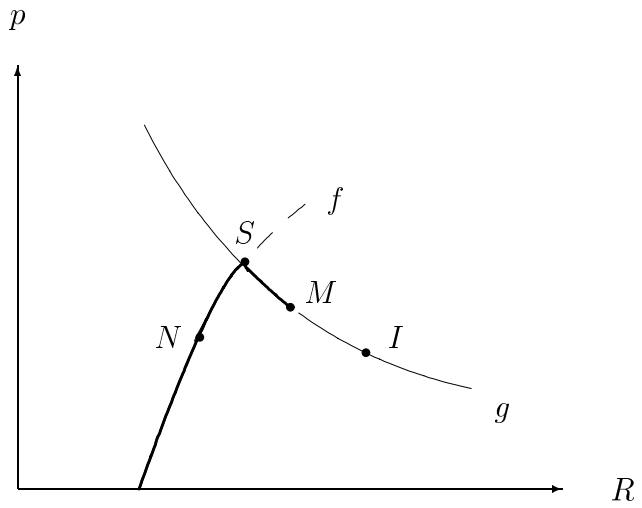

Figure 4.5: Comparing the free entry monopoly equilibrium M and the best competitive equilibrium C.

When the fixed costs are low, how does the resulting retail price-area pattern compare with that of the integrated optimum? Inspection of the derivative of the producer's profit along the f-curve ($\frac{d}{dR}\Pi_{prod}(f(R), R)$), evaluated at $p = p^I$, tells us that this derivative is positive. In other words, the optimal competitive equilibrium involves too small retail areas from the viewpoint of the producer (and the vertical structure). Potentially, several effects are at play. First, there is the business stealing effect: retailers enter without taking into account the effect on the other retailers in terms of reduced profits. An integrated firm would internalize this externality and reduce the number of retail outlets. On the other hand, retailers do not take into account the positive externalities on the manufacturer either: more entry leads to a lower mark-up and, hence, to an alleviated double marginalisation problem. Here, an integrated firm could solve this problem independently of the number of outlets, viz. by the pricing instrument. Finally, there can be a marketing effect: retailers being at a smaller distance increases the scope for a better capture of consumer surplus. This effect is present in the current model, but not really in the competitive equilibria as such. The reason is that with a fixed transport cost structure, the amount that each consumer buys does not depend on the distance to the shop; only the decision whether to come to the shop depends on it. Given that in each competitive equilibrium, the

indifferent consumer derives positive utility from going to a shop, decreasing the distance to be covered by consumers does not bring extra demand. Neither does it allow for a better capture of consumers' surplus (at least, not marginally speaking; for contemplated large price increases things may be different). The marketing effect principally plays a role where consumers are on the edge of buying or not buying, in particular in the kinked equilibria and the free entry monopoly equilibrium. However, for the free entry monopoly equilibrium (and the kinked equilibria) we already saw that it involves smaller retail areas than optimal for the vertical structure as a whole.

The conclusion of this section is that the best possible non-integrated equilibrium (be it a competitive or a monopoly equilibrium) involves too high prices and too high an outlet density from the perspective of the vertical structure.

4.3.3 Price restraints

In general, the wholesale price will not suffice to generate the preferred price-area combination (p^I, R^I): the producer will have to rely on other instruments as well. In this section we will see whether price restraints (resale price maintenance, price floors, price ceilings) vis-à-vis the retailers can do the job.

If the producer is in a position to set the retail price itself, then the only equilibrium condition for the retail market is the zero-profit condition (**??**). For a specific choice of p, the zero-profit condition defines retail area R as a function of wholesale price p_w:

$$R(p, p_w) = \frac{e^p \cdot F}{(p - p_w) \cdot v}$$

This function has range $[\frac{e^p \cdot F}{(p-c)\cdot v}, \infty)$, given that wholesale prices will not be taken smaller than c (in the absence of franchise fees, pricing below marginal cost would result in a loss for the producer). This implies that, once the retail price p is set, any equilibrium distance R that is larger than $\frac{e^p \cdot F}{(p-c)\cdot v}$ can be obtained by taking the appropriate wholesale price. Hence, *any* point in the price-area space (as long as $R \geq \frac{e^p \cdot F}{(p-c)\cdot v}$) can be attained using the price and wholesale price instruments.

As we have seen, the best equilibrium that the producer can choose in the absence of vertical restraints involves too high prices and too high an outlet density from the perspective of the vertical structure. A price ceiling (at p^I) with accompanying wholesale price

$$p_w = p^I - \frac{e^{p^I} \cdot F}{R^I \cdot v}$$

is therefore a sufficient instrument to bring about the equilibrium (p^I, R^I). Note that whenever $F < F_{viable}$, we have $p_w > c$. The use of this vertical restraint brings about a reduction in prices $(p^I < p^N, p^M)$ and in outlet density $(R^I > R^N, R^M)$.

4.3.4 Two-part tariffs

The model also provides a useful framework to study the role of two-part tariffs (franchise fee plus wholesale price). Imposing a franchise fee amounts to increasing the fixed costs that each retailer faces and has the effect of shifting outward curve f and shifting the kinked equilibria and the free entry monopoly equilibrium along curve g.

One way to obtain the first best by means of a franchise fee is to insure that the fixed cost perceived by the retailer (the fixed cost F plus the franchise fee A) leads to a free entry monopoly equilibrium that coincides with I. For which value of the perceived fixed cost $(F + A)$ is it the case that the free entry monopoly equilibrium M coincides with the integrated optimum I? Inserting the retail area R^M we can calculate the value of $\frac{d}{dR}\Pi_{int}(g(R), R)$ in the free entry monopoly equilibrium:

$$\frac{d\Pi_{int}(g(R^M), R^M)}{dR} = \frac{1}{2}vtL(\ln\sqrt{\frac{v}{(F+A)t}} - c - 1) + \frac{F}{\frac{4(F+A)}{vt}}$$

The level of the franchise fee for which this derivative is zero is the level that leads to the industry optimum (in conjunction with wholesale price $p_w = \ln(\sqrt{v/(F+A)t}) - \frac{1}{2}$). In fact, one could equally choose one of the kinked equilibria using an appropriate franchise fee and wholesale price.

Like a price ceiling, a franchise fee has the effect of reducing the number of retailers entering in equilibrium. The saved entry costs accrue to the manufacturer.

4.4 Welfare assessment

In determining the welfare effects of the vertical restraints we take welfare to be the sum of producer's surplus (profits) and consumers' surplus:

$$W(p, R) = \pi(p, R) + CS(p, R)$$

Consumers' surplus $CS(p, R)$ can be written as n times the surplus of the consumers that visit a particular shop. The latter is

$$\begin{aligned} CS(p,R) &= 2v \cdot \int^{\frac{1}{2}R} (\cdot e^{-p} - tx)dx \\ &= R \cdot (v \cdot e^{-p} - \frac{1}{4}vtR) \end{aligned}$$

Realizing that, in equilibrium, $n = \frac{L}{R}$ and using (4.6), we obtain

$$\begin{aligned} W(p,R) &= Lv \cdot e^{-p} \cdot (p-c) - \frac{L}{R} \cdot F + L \cdot (v \cdot e^{-p} - \frac{1}{4}vtR) \\ &= L \cdot [v \cdot e^{-p} \cdot (p-c+1) - \frac{F}{R} - \frac{1}{4}vtR] \end{aligned}$$

By taking the first derivatives w.r.t. p and R and checking whether $p < g(R)$, we find that in the welfare optimum W:

$$p^W = c$$

$$R^W = 2\sqrt{\frac{F}{vt}}$$

The fact that $p^W = c$ is a very familiar result. Note that, as can be expected, R^W is increasing in F and decreasing in v and in t. It is easily verified that in the integrated optimum I (the preferred situation of the producer), retail prices are too high and outlet density is too low from the welfare perspective. The first result is again very familiar; the second stems from the fact that the producer is only interested in whether or not consumers come to the shops, not in the distance they have to cover.

When the fixed costs of setting up a retail outlet (F) are large, the free entry monopoly equilibrium is the point of reference with which optimum I should be compared. By the use of the price ceiling, outlet density becomes lower (which is bad for welfare, since $R^W = R^M < R^I$), but prices come down as well (good for welfare). The overall welfare effect is ambiguous.

The same holds when fixed costs are low and the best competitive equilibrium is the proper point of reference. By using resale price maintenance a producer is able to decrease retail margins (good for welfare) but to decrease outlet density (undetermined welfare effect, since $R^C < R^W < R^I$).

4.5 Conclusion

This chapter analysed the role of resale price maintenance in sales contracts between a producer and its distributors, using a model of spatial competition in which the opportunity costs of shopping are distance-dependent (as in the usual Hotelling setting) but not quantity-dependent: for each visit to a shop, a consumer incurs a certain cost which is independent of the quantity bought. As in different settings (Mathewson and Winter (1983), Bittlingmayer (1983), Dixit (1983)), it was found that the optimal non-integrated equilibrium involves too high a retail price and too small retail areas, from the viewpoint of the manufacturer and the vertical structure as a whole.

As in Mathewson and Winter (1983) and Bittlingmayer (1983), the business stealing effect (retailers enter without taking into account the effect on the other retailers in terms of reduced profits) is determinative: an integrated firm internalizes this externality by reducing the number of retail outlets. The co-ordination problems that exist within the vertical structure (the double marginalisation problem and the co-ordination problem on market coverage) can either be tackled using a different pricing instrument (price ceilings, two-part tariffs) or are not strong enough to offset the business stealing effect. In fact, in the competitive equilibria, the market coverage effect is even hardly present. Given that in each competitive equilibrium, the indifferent consumer derives positive utility from going to a shop, decreasing the distance to be covered by consumers does not bring extra demand (in any case, not marginally speaking). This is probably a driving force behind the result that even though the retailing conditions in the competitive equilibria appear to be more competitive with the fixed transportation cost structure than with other transportation cost structures, this model does not reveal a rationale for the 'outlets hypothesis' either[11].

None of the models that have looked at the outlets hypothesis in a pure spatial dynamic framework has exhibited a rationale for it. Given the very stylized character of the models (a single product, symmetry assumptions) there is probably need for more analysis. A natural extension of the current model would be to consider multiproduct retailers. This line of research is particularly interesting as it is sometimes alleged that when a product is used as a *loss leader* (meaning that the product is priced below cost in order to attract consumers into the retail premises), the size of the distribution network is put under pressure as the number of retailers willing to carry the product

[11] An analysis of different demand specifications, with different demand elasticities ($q(p) = e^{-ap}$, $q(p) = a - p$ or $q(p) = p^{-\epsilon}$ $(a, \epsilon > 0)$) has yielded the same results.

concerned reduces. It is, however, not obvious that a tractable model can be found to study this case as well; existing spatial models for multiproduct competition have run into serious technical difficulties (Stahl, 1996).

Another possible course of action is empirical research. Empirical research on the role of minimum resale prices is relatively scarce. A notable exception are Ippolito and Overstreet (1996) who have assessed the impact of a Federal Trade Commission's decision that effectively put an end to the policy of price floors used by Corning Glass Works, a U.S. manufacturer of kitchenware. They conclude that in the case of Corning the outlets theory seems a more likely explanation of its past policy of resale price maintenance than the other pro- and anti-competitive theories. In order to obtain more insight into the several hypotheses, it is clear that more empirical work would be very welcome.

4.6 Appendices

Proof of Lemma 4.1: competitive equilibria

The first order derivative of the profit maximization problem in a competitive equilibrium is:

$$\begin{aligned}\frac{\partial}{\partial p_i}\pi^c(p_i, p_{-i,}R) &= \frac{v}{t}e^{-p_i}[-e^{-p_i}\cdot(2p_i - 2p_w - 1) + e^{-p_i}\cdot(p_i - p_w - 1) \\ &\quad -tR\cdot(p_i - p_w - 1)]\end{aligned}$$

Putting it to zero and assuming symmetry in prices, $p_i = p_{-i} = p$, gives us the first order condition

$$(p - p_w)\cdot(e^{-p} + tR) - tR = 0 \tag{4.8}$$

The second order derivative equals

$$\begin{aligned}\frac{\partial^2}{\partial p_i^2}\pi^c(p_i, p_{-i,}R) &= -\frac{v}{t}e^{-p_i}[-e^{-p_i}\cdot(2p_i - 2p_w - 1) + e^{-p_i}\cdot(p_i - p_w - 1) - \\ &\quad -tR\cdot(p_i - p_w - 1)] + \frac{v}{t}e^{-p_i}[-e^{-p_i}\cdot(2p_i - 2p_w - 1) - \\ &\quad -e^{-p_i} + e^{-p_i} - tR]\end{aligned}$$

Let us evaluate this expression for relation f: $f(R) = \ln\frac{vtR^2 - F}{FtR}$. As $p = f(R)$ is positive we have $vtR^2 - F > FtR$, i.e. $R > \frac{1}{2v}(F + \sqrt{F^2 + 4vF/t})$. On relation f it holds that

$$\begin{aligned}p_{-i} &= f(R) \\ p_w &= f(R) - \frac{vtR^2 - F}{vtR^2}\end{aligned}$$

This enables us to write the second order derivative on f as

$$\begin{aligned}\frac{\partial^2}{\partial p_i^2}\pi^c(p_i, f(R), R) &= \frac{v}{t}e^{-p_i}\cdot[4e^{-p_i}\cdot(p_i - p_w - 1) + \\ &\quad tR\cdot\frac{vtR^2 - 2F}{vtR^2 - F}\cdot(p_i - p_w - 2)]\end{aligned} \tag{4.9}$$

Locally, at $p_i = f(R)$, the second order condition for profit maximization is satisfied:

$$\frac{\partial^2}{\partial p_i^2}\pi^c(f(R), f(R), R) = -\frac{FtR}{vtR^2 - F}[(vtR - F)^2 + vtR^2 + F] < 0$$

(as $f(R) > 0$, we have $vtR^2 - F > 0$). Locally, therefore, the second order condition to the profit maximization problem is satisfied by prices and market areas satisfying the first order condition (4.1). In fact, a sufficient condition for profit function $\pi^c(.)$ to be concave on the entire relevant price interval is that $R > \sqrt{2F/vt}$, since an upperbound on the relevant prices is $p_w + 1$, the equilibrium price under Löschian spatial competition[12] for given wholesale price p_w; therefore, for all relevant prices $p_i - p_w - 1 < 0$ and $p_i - p_w - 2 < 0$ in expression (4.9). Expression (4.9) is not conclusive on the concavity of the profit function when $R < \sqrt{2F/vt}$. However, in the next sections it will be shown that the case $R < \sqrt{2F/vt}$ can be disregarded: it implies low price levels $p = f(R)$, levels that turn out not to be optimal from the viewpoint of the manufacturer (or the vertical structure).

The above calculations have relied on the situation being 'competitive', i.e. retailers are competing with their neighbours for the indifferent consumer located in-between. Apart from marginally increasing or decreasing its price, each retailer also has the possibility of 'drastically' increasing or decreasing its price. Concerning the first possibility: let us check whether a retailer can increase its equilibrium profits (i.e. obtain a profit greater than zero), by raising its price by such an extent that the relevant market frontier is no longer determined by the indifferent consumer $\bar{x}$ but by location $x_{\max}$. In other words, can a retailer get positive profits by picking a point on its monopoly segment? A first observation is that the retailer, in order to arrive on its monopoly segment, will have to pick a price which is at least as high as the price associated with the kink in the demand curve. As profits on the monopoly segment are a concave function of the price, a retailer would either choose the monopoly price $p^m(p_w)$ or, if this price is not on the monopoly segment, the lowest price which is, i.e. the price associated with the kink. The latter price, p', is determined by

$$x_{\max}(p') = R - x_{\max}(p)$$

where p is the price of the other retailers. Using the relations $p = f(R)$ and $x_{\max}(p) = \frac{1}{t}e^{-p}$, one finds

[12] Under Löschian spatial competition, each outlet assumes that its market area is invariant to changes in its price. In other words, each outlet assumes that neighbouring outlets will match any price change. This amounts to maximizing $vR \cdot e^{-p}(p - p_w)$, which gives $p = p_w + 1$. See also Mathewson and Winter (1983), Capozza and Van Order (1977) or Lösch (1938).

$$p' = f(R) + \ln(\frac{F}{vtR^2 - 2F})$$

Note that a kink price only exists when $R > \sqrt{2F/vt}$. (when $R < \sqrt{2F/vt}$, we have that $R < x_{\max}(p)$, so that the location of each retailer is in fact included in the potential market area of the competitors: no monopoly segment of the demand curve exists in such a case). As the kink price 'connects' the competitive segment with the monopoly segment of each firm's demand curve and we have already established global concavity of the competitive segment for $R > \sqrt{2F/vt}$, we can conclude that choosing the kink price is no profitable option. Let us consider choosing the monopoly price. In a free entry monopoly equilibrium, retailers turn out to be not directly competing with each other. Profits for a local monopolist if it sets price p are

$$\pi_i^m(p) = q^m(p) \cdot (p - p_w) - F$$

where $q^m(p) = \frac{2v}{t} \cdot e^{-2p}$. Profit maximization gives[13]:

$$p^m(p_w) = p_w + \frac{1}{2}$$

On relation f it holds that $p^m(p_w) = p^m(f(R) - \frac{vtR^2 - F}{vtR^2}) = f(R) - \frac{vtR^2 - F}{vtR^2} + \frac{1}{2} = f(R) - \frac{1}{2} + \frac{F}{vtR^2}$. This price only exceeds $p = f(R)$ when $\frac{1}{2} + \frac{F}{vtR^2} > 0$, i.e. when $R < \sqrt{2F/vt}$. In other words, whenever a monopoly segment exists ($R > \sqrt{2F/vt}$), the price that an unconstrained local monopolist would like to set ($p^m(p_w)$) is not part of this segment (on relation f, price $p^m(p_w)$ gives rise to overlapping, not disjoint potential market areas; cf. Salop (1979; figure 6)). The best the firm can then do is to choose the lowest price which is on the monopoly segment, i.e. the price associated with the kink, but this price does not give (as outlined above) a higher profit than the best competitive price.

Let us now consider 'drastic' price cuts, in particular 'mill price undercutting': setting the price so low that the competitor's (effective) price at its own retail outlet is undercut. At this price level there is a discontinuity in the demand curve: for prices slightly above the level, the market area is defined by the indifferent consumer located inbetween; for prices slightly below, the entire market area previously served by the competitor is now served by the undercutting firm (cf. Eaton and Lipsey (1978), d'Aspremont e.a. (1979) or Novshek (1980)). Retailer i undercuts the mill price of retailer j when $\bar{x}(p) = R$:

[13] See also the next appendix section.

$$\frac{1}{2t} \cdot (e^{-p_i} - e^{-p_j} + tR) = R \text{ , i.e. } \hat{p}_i = \ln(tR + e^{-p_j})^{-1}$$

Evaluated on $p_j = f(R)$, the mill price undercutting price is

$$\begin{aligned} \hat{p}_i &= -\ln(tR + \frac{FtR}{vtR^2 - F}) \\ &= \ln \frac{vtR^2 - F}{Ft^2R^3} \\ &= \ln \frac{vtR^2 - F}{FtR} - \ln \frac{vtR^2}{F} \\ &= f(R) + \ln \frac{F}{vtR^2} \end{aligned}$$

Profits are then

$$\begin{aligned} \pi^c(\hat{p}_i, f(R), R) &= v \cdot 3R \cdot e^{-\hat{p}_i}(\hat{p}_i - p_w) - F \\ &= v \cdot 3R \cdot e^{-f(R)} \cdot \frac{vtR^2}{F}(\ln \frac{F}{vtR^2} - \frac{F}{vtR^2} + 1) - F \end{aligned}$$

As $\frac{d}{dz}(\ln z - z + 1) = \frac{1}{z} - 1 > 0$ for $z < 1$ and $\frac{F}{vtR^2} < 1$, we conclude that $\ln \frac{F}{vtR^2} - \frac{F}{vtR^2} + 1 < 0$, and hence $\pi^c(\hat{p}_i, f(R), R) < 0$: mill price undercutting will not occur in the competitive equilibrium. A 'no mill price undercutting' assumption, as put forward by Eaton and Lipsey (1978) or Novshek (1980), is therefore not necessary. Note, however, that this is due to our assumption of symmetric locations: without this assumption, the pay-off function would not be defined for small differences in location in the first place.

We conclude that the combinations (p, R) on f, satisfying the requirement of compatibility with demand (g), represent competitive equilibria.

Proof of Lemma 4.2: the free entry monopoly equilibrium

In a free entry monopoly equilibrium, retailers turn out to be not directly competing with each other. Profits for a local monopolist if it sets price p_i are

$$\pi_i^m(p_i) = q^m(p_i) \cdot (p_i - p_w) - F$$

The first order derivative of the profit maximization problem in a monopoly equilibrium is:

$$\frac{\partial}{\partial p_i}\pi^m(p_i) = -\frac{4v}{t} \cdot e^{-2p_i} \cdot (p_i - p_w) + \frac{2v}{t} \cdot e^{-2p_i}$$

Putting it to zero gives us the FOC: $p^m(p_w) = p_w + \frac{1}{2}$. The second order derivative equals

$$\begin{aligned}\frac{\partial^2}{\partial p_i^2}\pi^m(p_i) &= -\frac{2v}{t}e^{-p_i}\cdot[-2e^{-2p_i}\cdot(2p_i - 2p_w - 1) + 2e^{-2p_i}] \\ &= -\frac{4v}{t}e^{-3p_i}\cdot[-p_i + 1 + p_w] < 0\end{aligned}$$

as $p_i < p_w + 1$. The second order condition is therefore satisfied, at least, for the prices for which the above demand function specification is appropriate. Let us also check whether the free entry monopoly equilibrium can be subject to mill price undercutting. As indicated in the text, the price at the free entry monopoly equilibrium is $p^M = \ln(\sqrt{v/Ft})$, the market area is $R^M = 2\sqrt{F/vt}$ and the underlying wholesale price is $p_w^M = \ln(\sqrt{v/Ft}) - \frac{1}{2}$. The mill price undercutting price in the monopoly equilibrium is

$$\begin{aligned}\hat{p}_i &= \ln(tR + e^{-p_j})^{-1} \\ &= \ln\frac{1}{3}\sqrt{\frac{v}{Ft}}\end{aligned}$$

Profits for the undercutting firm are then

$$\begin{aligned}\pi(\hat{p}_i) &= v\cdot 3R^M\cdot e^{-\hat{p}_i}(\hat{p}_i - p_w) - F \\ &= v\cdot 3\cdot 2\sqrt{\frac{F}{vt}}\cdot 3\sqrt{\frac{Ft}{v}}\cdot(\ln\frac{1}{3}\sqrt{\frac{v}{Ft}} - \ln\sqrt{\frac{v}{Ft}} + \frac{1}{2}) - F \\ &= 18F(\ln\frac{1}{3} + \frac{4}{9}) < 0\end{aligned}$$

We conclude that the free entry monopoly equilibrium is not subject to mill price undercutting.

Proof of Lemma 4.3: the kinked equilibria

In the third type of equilibria, the potential market areas just touch each other and the profit maximizing prices are corner solutions. Each wholesale price $p_w \in [p_w^S, p_w^M)$ gives rise to one equilibrium price that satisfies $p = g(R)$. Let us check whether mill price

undercutting can occur. The mill price undercutting price in a kinked equilibrium is

$$\begin{aligned}\hat{p}_i &= \ln(tR + e^{-p})^{-1} \\ &= \ln \frac{2}{3tR}\end{aligned}$$

The wholesale price associated with a kinked equilibrium follows from the zero profit condition $\pi^k(p, R) = R \cdot v \cdot e^{-p}(p - p_w) - F = 0$:

$$p_w = \ln \frac{2}{tR} - \frac{2F}{vtR^2}$$

Profits for a mill price undercutting firm are therefore

$$\begin{aligned}\pi(\hat{p}_i, p, R) &= v \cdot 3R^k \cdot e^{-\hat{p}_i}(\hat{p}_i - p_w) - F \\ &= v \cdot 3 \cdot R \cdot 3\frac{tR}{2} \cdot (\ln \frac{2}{3tR} - \ln \frac{2}{tR} + \frac{2F}{vtR^2}) - F \\ &= \frac{9}{2}vtR^2(\ln \frac{1}{3} + \frac{2F}{vtR^2}) - F < 0\end{aligned}$$

as for the kinked equilibria $R > R^S = \sqrt{3F/vt}$, so that $\ln \frac{1}{3} + 2F/vtR^2 < \ln \frac{1}{3} + \frac{2}{3} < 0$. No mill price undercutting will occur at the kinked equilibria.

Proof of Propositions 4.1 and 4.2: the non-integrated optimum

Recall from Section 4.2 that three possible equilibria may arise. For low wholesale prices p_w, the resulting equilibria are the competitive equilibria, for intermediate p_w the resulting equilibria are the kinked equilibria and for $p_w = p_w^M$ the free entry monopoly equilibrium will be the result.

Let us first consider the kinked equilibria and the free entry monopoly equilibrium. They resemble the integrated optimum (p^I, R^I) in that the customers who are located furthest away from a shop are left with exactly zero utility (it holds that $p = g(R)$). It follows immediately that the free entry monopoly equilibrium dominates the kinked equilibria form the viewpoint of the manufacturer: for all kinked equilibria $p^k > p^M$ (so that demand is lower than in the free entry monopoly equilibrium), whereas $p_w^k < p_w^M$ (the wholesale margin for the manufacturer is lower as well).

Let us now compare the integrated optimum (p^I, R^I) with the monopoly equilibrium. It is unfortunate that the explicit coordinates of the integrated optimum I are not known to us, but this is not problematic. Inspection of $\Pi_{int}(g(R), R)$ reveals that along curve g profit Π_{int} is increasing for $p < p^I$ and decreasing for $p > p^I$. As we

know the coordinates of the free entry monopoly equilibrium M, we can calculate the value of $\frac{\partial}{\partial R}\Pi_{int}(g(R),R)$ in M:

$$\begin{aligned}\frac{\partial \Pi_{int}(g(R^M),R^M)}{\partial R} &= \frac{1}{2}vLt[(\ln\sqrt{\frac{v}{Ft}}-c-1)+\frac{1}{2}] \\ &= \frac{1}{2}vLt(p^M-c-\frac{1}{2})\geq 0\end{aligned}$$

as it holds that $p^M = p^m(p_w^M) \geq p^m(c) = c+\frac{1}{2}$ for every $p_w \geq c$. So, unless $p_w^M = c+\frac{1}{2}$, the monopoly equilibrium M does not coincide with the industry optimum I; M lies 'north-west' of I along the g-curve. It confirms that none of the kinked equilibria coincides with I and that none of them yield more profit to the producer than the free entry monopoly equilibrium M (for all kinked equilibria $R^k < R^M < R^I$).

The derivative of $\frac{\partial}{\partial R}\Pi_{int}(g(R),R)$ in M is zero (M and I coincide) if and only if $p_w^M = c$, i.e. the highest possible wholesale price equals marginal cost. This will be the case when the fixed costs are that large that even a vertically integrated firm would only just be able to break even in its operations. The level of fixed costs for which this is case is $F_{viable} = \frac{v}{t}e^{-2c+1}$. The conclusion is that, unless $F = F_{viable}$, the free entry monopoly equilibrium always involves a retail price and an outlet density that is too high from the viewpoint of the integrated structure.

What remains is to compare the overall profits in free entry monopoly equilibrium M with those of the best possible competitive equilibrium, C. Equilibrium C is found by maximizing $\Pi_{prod}(f(R),R)$ subject to $p \leq g(R)$. The derivative of Π_{prod} w.r.t. R along relation f:

$$\begin{aligned}\frac{\partial \Pi_{prod}(f(R),R)}{\partial R} &= -vL\cdot e^{-f(R)}\cdot f'(R)\cdot(f(R)-c-1)+\frac{LF}{R^2} \\ &= \frac{L}{R}F\cdot[-\frac{vFtR(vtR^2+F)}{(vtR^2-F)^2}\cdot(\ln\frac{vtR^2-F}{FtR}-c-1)+ \quad (4.10)\end{aligned}$$

An interior optimum (where the above derivative is zero) is only implicitly defined. A sufficient condition to have a corner solution (Π_{prod} increasing in R all along relation f) is that $f(R) \geq c+1$. Then the best competitive equilibrium is given by point S, where $f(R) = g(R)$. But in this case, we already know that the manufacturer prefers the monopoly equilibrium ($\Pi_{prod}(g(R),R)$ increases in R until $R = R^I$). Hence, a sufficient condition for the free entry monopoly equilibrium to be the non-integrated optimum is that $f(R) \geq c+1$, i.e. that $F \geq \frac{4v}{3t}e^{-2(c+1)}$. Note that $\frac{4v}{3t}e^{-2(c+1)} \leq F_{viable}$,

so there exist values of F that satisfy the sufficient condition.

Let us see whether it is possible that the best competitive equilibrium outperforms the free entry monopoly equilibrium for small F. This indeed appears to be the case. Profits in the free entry monopoly equilibrium are

$$\begin{aligned}\Pi_{prod}(p^M, R^M) &= L \cdot v \cdot e^{-p^M} \cdot (p^M - c) - \frac{L}{R^M} F \\ &= L \cdot \sqrt{Fvt}(\ln \sqrt{\frac{v}{Ft}} - \frac{1}{2} - c)\end{aligned}$$

Taking the limit for $F \to 0$ gives

$$\lim_{F\to 0} \Pi_{prod}(p^M, R^M) = \lim_{F\to 0} \frac{\frac{Lvt}{2\sqrt{Fvt}}}{(\sqrt{\frac{v}{Ft}})^{-1} \cdot (2\sqrt{\frac{Ft}{v}})^{-1} \cdot (-\frac{v}{tF^2})} = \lim_{F\to 0} \frac{Lt\sqrt{t}F}{\sqrt{v}} = 0$$

(using the theorem of de l'Hopital). The reason is that when $F \approx 0$, in order to have the retailers obtain a zero profit in a monopoly situation requires ever higher wholesale prices p_w. The resulting retail prices are then way above the optimal level, p^I. In fact, both p^M and p_w^M go towards infinity, leaving zero profit. When the fixed costs of setting up a retail outlet are almost zero, the free entry monopoly equilibrium cannot possibly be optimal. The optimum will be a competitive equilibrium, one implied by the first order condition based on (4.10). For example, when $F \approx 0$, the first order condition on profit maximization (4.1) and the zero profit condition (4.2) imply $p \approx p_w$ and $R \approx 0$, so that $\Pi_{prod}(p, R) \approx L \cdot v \cdot e^{-p_w}(p_w - c)$. The latter is maximized for $p_w \approx c + 1$ and gives a profit $\Pi_{prod} \approx L \cdot v \cdot e^{-c-1} > 0$.

Supposing that the fixed costs are such that the optimum for the manufacturer in the absence of vertical restraints is a competitive equilibrium, how does the resulting retail price compare with that of the integrated optimum? For this we only have to consider the sign of $\frac{\partial}{\partial R}\Pi_{prod}(f(R), R)$ evaluated at $p = p^I$. Recall from (4.2) that p^I satisfies

$$\frac{1}{2}tv(p - c - 1) + \frac{F}{(R^I)^2} = 0$$

This gives:

$$\frac{\partial}{\partial R}\Pi_{prod}(f(R), R)\Big|_{p^I} = -vL \cdot e^{-p^I} \cdot f'(R)|_{p^I} \cdot (-\frac{2F}{vt(R^I)^2}) + \frac{LF}{R^2} > 0$$

What we can conclude, is that the best possible non-integrated equilibrium (be it a competitive or a monopoly equilbrium) involves too high prices and too high outlet density from the perspective of the vertical structure as a whole.

4.7 References

d'Aspremont, C., Gabszewicz, J. and F. Thisse (1979) 'On Hotelling's stability in competition' Econometrica, 47: 1145-1150.

Bittlingmayer, G. (1983) 'A model of vertical restriction and equilibrium in retailing' Journal of Business 56(4): 477-498.

Capozza, D. and Van Order, R. (1977) 'Pricing under spatial competition and spatial monopoly' Econometrica 45(6): 1329-1338.

Dixit, A. (1983) 'Vertical integration in a monopolistically competitive industry' International Journal of Industrial Organization 1(1): 63-78.

Eaton, B. and Lipsey, R. (1978) 'Freedom of entry and the existence of pure profit' Economic Journal 88: 455-469.

Gould, J. and Preston, L. (1965) 'Resale price maintenance and retail outlets'. Economica 32: 303-12.

Holton, R.H. (1957) 'Price discrimination of retail: the supermarket case' Journal of Industrial Economics 6(1): 13-32.

Ippolito, P. and Overstreet, T. (1996) 'Resale price maintenance: an economic assessment of the Federal Trade Commission's case against the Corning Glass Works' Journal of Law and Economics XXXIX: 285-328.

Mathewson, G. and Winter, R. (1983) 'Vertical integration by contractual restraints in spatial markets' Journal of Business 56(4): 497-526.

Mathewson, G. and Winter, R. (1984) 'An economic theory of vertical restraints' Rand Journal of Economics 15: 27-38.

Novshek, W. (1980) 'Equilibrium in simple spatial (or differentiated product) models' Journal of Economic Theory 22: 313-326.

Reagan, P. (1986) 'Resale price maintenance: a re-examination of the outlets hy-

pothesis' Research in Law and Economics 9: 1-12.

Salop, S. (1979) 'Monopolistic competition with outside goods' Bell Journal of Economics 10: 141-156.

Stahl, K. (1982) 'Location and spatial pricing theory with nonconvex transportation cost schedules' The Bell Journal of Economics 13(2): 575-582.

Stahl, K. (1996) 'Towards a microeconomic theory of the retailing sector' mimeo, University of Mannheim.

Tirole, J. (1988) 'The theory of industrial organization' MIT Press, Cambridge MA, USA.

Vickrey, W. (1964) 'Microstatics' Harcourt, Brace and World, New York, USA. Reproduced in International Journal of Industrial Organization 17 (1999) 953-963.

Yamey, B.S. (1954) 'The economics of resale price maintenance' London: Sir Isaac Pitman.

Chapter 5

Resale price maintenance under cost uncertainty: a note on 'The logic of vertical restraints'

5.1 Introduction

As discussed in Chapters 2 and 4, several explanations for the use of resale price maintenance have been given in the economic literature. The focus of the present chapter is on the incentive and insurance properties of resale price maintenance in an uncertain trading environment, a subject that has notably been analysed by Rey and Tirole in their insightful contribution 'The logic of vertical restraints' (1986, American Economic Review 76: p. 921-936).

Rey and Tirole (1986) set up a spatial model of retail competition and analyse the role of resale price maintenance (RPM) when there is uncertainty about future demand and cost levels. The retailers are better informed (ex post) about the realisation of final demand and about their own costs than the manufacturer. The latter is assumed not to be able to obtain this information, directly or indirectly (one of the reasons for this is that the retailers can engage in arbitrage, making it impossible for the manufacturer to follow track of the supplies). Hence, informational problems prevent the manufacturer from using explicit contracts based on the true performances (profits) of the retailers.

The basic trade-off in the choice of contract is between the optimal exploitation of market power and the amount of risk that the retailers are willing to accept. The opti-

mal exploitation of market power requires that one avoids the double marginalisation problem associated with linear wholesale pricing, i.e. the problem that final prices end up too high from the viewpoint of the vertical structure due to consequent stages of market power (Spengler, 1950). One way of doing so is to use two-part tariffs, involving a fixed upfront payment (the franchise fee) and a low marginal wholesale price. Another method is resale price maintenance, by which the producer directly imposes the proper final price on the retailers. If there is no uncertainty, the two methods yield identical results. However, when there is uncertainty the two may differ. Whereas free competition between retailers clearly allows better use of local information than the inflexible instrument of RPM, the insurance properties are more complex: risk averse retailers dislike the possibility of not earning back a franchise fee, but also RPM exposes them to risks, to the extent that their profit margins are subject to fluctuations. The relative profitability of the two instruments depends on the circumstances in the market[1].

One of Rey and Tirole's (1986) more specific results is that when uncertainty is about *costs* and retailers are risk averse, free competition between retailers (in combination with a two-part tariff) has good incentive and insurance properties. It turns out to be a more profitable option for the manufacturer than RPM in the three model specifications analysed in the article:

1. the case in which cost uncertainty is market-wide (the costs faced by all retailers fluctuate in the same way) and retailers are non-differentiated
2. the case in which cost uncertainty is market-wide and retailers are differentiated[2] and
3. the case in which cost uncertainty is firm-specific ('idiosyncratic') and retailers are non-differentiated.

The principal reason for these results is that under cost uncertainty, the retail margin is particularly volatile under RPM: the retail price is fixed but the cost level varies, a feature for which risk averse retailers need some compensation (e.g. via lower wholesale prices).

[1]Rey and Tirole (1986) also consider exclusive distribution arrangements. The focus of this chapter is on the comparison between competition and RPM.

[2]Retailers being differentiated amounts to retailers posessing some (localized) market power: when, for example, 'shopping costs' are non-zero, proximity differentiates the retailers from the viewpoint of the customers.

The main goal of the current chapter is to show that the result on the favourable incentive and insurance properties of competition does not generally carry over to the case of firm-specific cost uncertainty and retailers being differentiated (i.e. possessing some market power).

The essential point is that in the absence of retailer differentiation (as in the case of idiosyncratic cost uncertainty in Rey and Tirole (1986)), there is only one source of double marginalisation: the double marginalisation due to cost differences at the retail level (when one retailer turns out to be more efficient than the other, it obtains a positive retail margin by just undercutting the price of the other retailer). With only this source of double marginalisation, the need for a powerful two-part wholesale tariff is not very great: even when retailers are so risk averse that they are not willing to accept any positive franchise fee at all (because this might result in a loss in some situations), downstream competition turns out to perform better than resale price maintenance[3].

However, when there is also a double marginalisation problem due to retailer differentiation, there is a greater need for a powerful wholesale tariff, involving lower wholesale prices and, correspondingly, a higher franchise fee. But the extent to which the manufacturer can charge the required franchise fee is limited by the risk that the retailers are willing to bear: a franchise fee can only be recovered by the retailers when the retail margin they earn is sufficiently positive (in some average sense)[4]. Whereas firm-specific cost uncertainty makes it more difficult to charge a franchise fee to risk averse retailers which are competing in prices, resale price maintenance is an instrument to protect the retailers against more efficient rivals. As a result of this insurance property, resale price maintenance can be an optimal commercial policy for a manufacturer. In particular, it is shown that in the case of differentiated retailers, RPM is optimal when the cost uncertainty is firm-specific, the retailers are sufficiently risk averse and the range of possible retail cost levels is not too wide (so as to make RPM too 'rigid' as an instrument).

The next section, Section 5.2, will describe the model; it involves a specification of retail differentiation that is simpler than that of Rey and Tirole (1986), in order to be

[3]Cf. Rey and Tirole (1986), p. 930. When no franchise fee is imposed, the insurance properties of free retail competition are necessarily very good: even when the retail margin ends up being zero (which occurs whenever the competitors are at least as cost efficient), the retailers obtain a net profit equal to zero.

[4]In a similar fashion, the results of Rey and Tirole (1986) also depend on the specific assumption made as regards the level of fixed costs required to set up a retail outlet.

able to conduct the analysis. Section 5.2 also characterises the retail market equilibria for given wholesale prices and the industry optimum. Section 5.3 addresses the optimal choice of contract from the viewpoint of the manufacturer, when retailers are risk neutral; Section 5.4 does the same for the case of risk averse retailers. Concluding comments and a short discussion are offered in Section 5.5.

5.2 The basic model

As indicated in the introduction, the role of resale price maintenance will be studied in a simple model of retail competition. We assume that a manufacturer supplies its product to (at most) two retailers. The unit costs of production equal c. The retailers are differentiated from the viewpoint of the consumers. If the product is sold by two retailers, market demand d_i for retailer i $(i = 1, 2)$ depends negatively on the own price p_i and positively on the price of the competing retailer, p_j $(j \neq i)$:

$$d_i(p_i, p_j) = 1 - p_i + \beta \cdot p_j \qquad (0 \leq \beta < 1)$$

The above demand specification[5] differs from the spatial competition framework of Rey and Tirole (1986). This is done in order to keep the analysis tractable, in particular in order to obtain closed form solutions for the equilibrium price levels and profit levels in case the retail costs of the retailers are not the same (this appeared impossible in a spatial competition framework with downward sloping consumer demand).

Each retailer that decides to resell the product incurs a fixed investment cost $F \geq 0$[6]. The marginal retailing costs of retailer i are denoted by γ_i. In line with Rey and Tirole (1986), the cost parameters are, ex ante, not known to the retailers or the manufacturer: it is only common knowledge that they are independently distributed following a uniform distribution: $\gamma_i, \gamma_j \sim U(\underline{\gamma}, \bar{\gamma})$, where $0 \leq \underline{\gamma} \leq \bar{\gamma} < \frac{1}{1-\beta} - c$. The true values (the 'realisations') are observed by the retailers after they have entered the market. Subsequently, they can use this information to make their pricing decisions. By contrast, the cost parameters cannot be observed or inferred by the manufacturer

[5] See Appendix 5.6.1 for the underlying utility function.

[6] A tempting interpretation of the fixed costs is that they represent the opportunity costs of shelf space (profits foregone by having the product concerned in the shelves and not some other product). However, this interpretation implies a multiproduct setting, a much more complex context (cf. Rey and Vergé (1999) for such an analysis in the absence of uncertainty).

(any obtained information by the manufacturer is at best 'soft')[7]. It is assumed that informational problems only allow for the use of a two-part wholesale tariff $T(.)$ (consisting of a fixed fee A and a wholesale price p_w) and/or the use of resale price maintenance. The contract offered by the manufacturer is public knowledge[8]. The manufacturer is taken to be risk-neutral, an assumption which can be justified by the assumption that the manufacturer supplies many, statistically independent markets.

Retailers will enter the market if and only if the utility obtained from the prospective profits is higher than, or equal to zero. As regards the risk attitudes of the retailers, we consider two extremes: risk neutral retailers and infinitely risk averse retailers. The first type is willing to enter the market and invest F as soon as the expected net profit is at least zero. The second type invests only if it is sure that the net profit is at least zero in every possible state of nature.

The exact timing of the model is as follows:

- stage I: the manufacturer decides on the (uniform) sales conditions vis-a-vis the retailers, i.e. on a wholesale price p_w, on a fixed fee (or subsidy) A, and on whether or not to use resale price maintenance;

- stage II: given the manufacturer's choice, retailers decide simultaneously whether or not to enter a market; If they enter, they invest an amount $F \geq 0$.

- stage III: the retailers which have entered a market learn the realisations of the previously uncertain cost parameters; the manufacturer, however, remains ignorant about the realized values of the parameters.

- stage IV: the retailers compete in prices (Bertrand competition between differentiated retailers). If only one retailer has entered, it can act as a monopolist.

In stage I, preceding the entry decision of the retailers, the producer decides on the sales conditions vis-à-vis the retailers with the objective of maximizing its profits. In doing so, the manufacturer takes into account how its decisions will affect the retailers' behaviour. As a preliminary step, I will characterise the free retail market equilibria that arise for given wholesale prices and given cost parameters. Then, I will characterise the 'first-best' result, i.e. the result that maximizes profit for the vertically integrated structure. This will serve as a benchmark with which the performance of the respective instruments (wholesale price, fixed fee, the vertical restraints) can be assessed.

[7] As mentioned in the introduction, one explanation for this is that the retailers can engage in arbitrage, making it impossible for the manufacturer to follow track of the supplies.

[8] No commitment problems in the sense of Rey and Tirole (1997) or Segal (1999) arise, therefore.

5.2.1 Characterisation of retail market equilibria for given wholesale prices and given cost parameters

Consider first the situation that two retailers have entered in stage II and that the cost realisations have been learnt in stage III. In stage IV, given that retailer j charges price p_j for the product it sells, the profits for retailer i are

$$\begin{aligned}\pi_i(p_i, p_j) &= (p_i - p_w - \gamma) \cdot d_i(p_i, p_j) - F \\ &= (p_i - p_w - \gamma) \cdot (1 - p_i + \beta \cdot p_j) - F\end{aligned}$$

The first order condition of profit maximization gives[9]

$$p_i = \frac{1}{2} \cdot (1 + \beta \cdot p_j + p_w + \gamma_i) =: R_i(p_j)$$

In the Nash equilibrium $p^* = (p_i^*, p_j^*)$, it holds that $p_i = R_i(R_j(p_i))$ or

$$p_i^*(p_w) = \frac{1}{2 - \frac{1}{2}\beta^2}(1 + p_w + \gamma_i + \frac{1}{2}\beta(1 + p_w + \gamma_j)) \tag{5.1}$$

where the retail equilibrium price p_i^* for retailer i $(i = 1, 2)$ is written as a function of the marginal wholesale price p_w charged by the manufacturer. The retailers' profits are, expressed as functions of the wholesale price p_w:

$$\pi_i^*(p_w) = \frac{1}{4(1 - \frac{1}{2}\beta)^2}(1 - (1 - \beta)(p_w + \gamma_i) + \frac{\frac{1}{2}\beta}{1 + \frac{1}{2}\beta}(\gamma_j - \gamma_i))^2 - F \tag{5.2}$$

When, in stage II, only one retailer has entered (say, retailer i), it can act as a monopolist in the pricing stage. It will then face the demand function[10]

$$d_i(p_i) = 1 + \beta - (1 - \beta^2) \cdot p_i$$

The corresponding monopoly price is

$$p_i^m(p_w) = \frac{1}{2}(p_w + \gamma_i) + \frac{1}{2(1 - \beta)} \tag{5.3}$$

and the profits are

$$\pi_i^m(p_w) = \frac{(1 + \beta)(1 - (1 - \beta)(p_w + \gamma_i))^2}{4(1 - \beta)} - F \tag{5.4}$$

[9] The second order condition is satisfied. It is assumed, for the moment, that the expression for $R_i(p_j)$ is at least equal to $p_w + \gamma_i$; this point will be adressed later on.

[10] See the appendix for its derivation from the underlying system of demand curves $d_i(p_i, p_j) = 1 - p_i + \beta \cdot p_j$ $(i = 1, 2)$ posited earlier.

5.2.2 Benchmark case: the vertically integrated structure

In stage I, preceding the entry decision of the retailers, the producer decides on the sales conditions vis-à-vis the retailers with the objective of maximizing its profits. As a first step, I shall characterise the 'first-best' result, i.e. the result that maximizes profit for the vertically integrated structure. This will serve as a benchmark with which the performance of the respective instruments (wholesale price, fixed fee, resale price maintenance) can be assessed. In the next section, the optimal choice for the manufacturer will be characterised, both for the case that retailers are risk-neutral and for the case that they are extremely risk averse.

A first question to answer is: how many outlets would a vertical structure want to have? One or two? The trade-off is as follows: on the one hand, the more shops the structure opens, the higher the fixed costs involved are. On the other hand, opening more shops offers more variety to consumers and, hence, more consumer demand. In comparing the maximally obtainable profits under both options, we obtain the following proposition (where we have assumed that the vertical structure only cares about expected profits, just as the risk neutral manufacturer):

Proposition 5.1 *A sufficient condition for the vertically integrated firm to prefer two retail outlets instead of one is that the investment cost per retail outlet is not too large:* $F \leq \frac{1}{4} \cdot (1 - (1 - \beta)(c + \gamma^e))$, *where* $\gamma^e = \mathbf{E}\{\gamma_i\}$.

Proof. See appendix.

The fact that it is better to have two retail outlets instead of one when the cost of investment is not too large is intuitive. First of all, consumers value retail variety and to offer this variety means that the manufacturer can extract more surplus from the consumers. The possible uncertainty about costs only reinforces the result: for a given value of the expected marginal retail cost (γ^e), the expected profit for the integrated structure, both when there is one outlet and when there are two, is increasing in the spread ($\bar{\gamma} - \underline{\gamma}$). More precisely, the profit functions are convex in $\bar{\gamma}$ and $\underline{\gamma}$: a drop in profits following a rise in retail cost is limited as the retail price can be adjusted optimally, limiting the drop in sales. This holds even more for the case of two firms, because it entails one more degree of freedom.

In what follows, we will focus on cases in which it is optimal for the industry to have two retail outlets, as only in these cases there is an issue of idiosyncratic risk to be studied. When the structure chooses for two retail outlets it maximizes

$$\pi^I_{two}(p_i, p_j) = \sum_{i=1, j\neq i}^{2} (p_i - c - \gamma_i) \cdot d_i(p_i, p_j) - 2F$$

The (ex post) optimal retail prices and their expected levels are stated, for future reference, in the following lemma:

Lemma 5.1 The (ex post) optimal retail prices for the integrated firm when it opts for two outlets are $p^I_i = \frac{1}{2(1-\beta)} \cdot (1 - (1-\beta)(c + \gamma_i))$, $(i = 1, 2)$. In expectation,

$$\mathbf{E}\{p^I_i\} = \frac{1}{2(1-\beta)} \cdot (1 - (1-\beta)(c + \gamma^e))$$

Proof. See appendix.

5.3 The manufacturer's choice: the case of risk neutral retailers

5.3.1 Two-part tariffs

If a producer cannot use vertical restraints, the only decision variable is the price (or price schedule) at which it sells to the retailers, the wholesale tariff. In this section, we will consider the outcomes that a two-part wholesale tariff can produce. If q is the quantity sold to the retailer, the latter pays wholesale tariff $T(q) = A + p_w \cdot q$, where A is the franchise fee and p_w the unit wholesale price. We will not explicitly consider the case of a traditional linear tariff as it is, first, just a special case of a two-part tariff (the franchise fee being zero) and, secondly, unlikely to bring about an optimal outcome, given the problem of double marginalisation: since the retailers are differentiated, there will be some double marginalisation and because consumer demand is elastic, it brings about the usual demand distortion.

Let us suppose that the producer aims at having two retailers carrying its product. This amounts to assuming that the fixed cost of investment is not too large (this assumption will be checked later in the appendix). When the retailers are risk neutral they are willing to accept the sales conditions of the manufacturer whenever it gives them an non-negative profit in expectation. So, the manufacturer chooses franchise fee A and unit wholesale price p_w so as to solve the following program:

$$\max_{A,p_w} \mathbf{E}\{\pi_{prod}(p_w)\} + 2 \cdot A$$
$$\text{s.t. } A \leq \mathbf{E}\{\pi_i(p_w)\}$$

where $\pi_{prod}(p_w)$ is the producer's operating profit (excluding the franchise fee) when it charges marginal wholesale price p_w,

$$\pi_{prod}(p_w) = (p_w - c) \cdot \sum_{i=1,2} d_i(p_1(p_w), p_2(p_w)),$$

and $\pi_i(p_w)$ is retailer i's profit (excluding the franchise fee) given by expression (5.2). The maximum franchise fee that the manufacturer can charge is such that the retailers are left without profit, in expectation. This maximal fee is also the optimal one: it is not optimal for the producer to leave profit in the hands of the retailers. Hence, the maximization problem reduces to

$$\max_{p_w} \mathbf{E}\{\pi_{prod}(p_w)\} + 2 \cdot \mathbf{E}\{\pi_i(p_w)\}$$

We state the following lemma:

Lemma 5.2 *(Risk neutral retailers).* With competing retailers, the optimal marginal wholesale price for the manufacturer is given by $p_w^A = c + \frac{1}{2}\beta(\frac{1}{1-\beta} - c - \gamma^e)$, the optimal franchise fee is $A = \mathbf{E}\{\pi_i(p_w^A)\}$. The manufacturer eliminates the double marginalisation problem in expectation ($\mathbf{E}\{p_i\} = \mathbf{E}\{p^I\}$) but, generically, not ex post. The manufacturer's expected profit does not attain that of the integrated optimum.

Proof. See appendix.

Note that, in this case, the optimal marginal wholesale price is larger than the marginal cost of production, c: the wholesale price is, as it should be, chosen to reflect the (expected) opportunity cost of selling one more unit to one of the retailers. After all, if one more unit is sold to a retailer, there is not only a marginal cost of production c to the structure, but also the effect that every extra unit sold by a retailer depresses the price that the other retailer can charge to the consumers. In fact, the term $\frac{1}{2}\beta(\frac{1}{1-\beta} - c - \gamma^e)$ in Lemma 5.2 represents the partial derivative of the expected profit of retailer i with respect to the quantity sold by retailer j [11]. In expectation, the potential double marginalisation problem associated with retail market power is avoided by taking a marginal wholesale price that gives just the

[11] The partial derivative of the expected profit of retailer i with respect to the quantity sold by retailer j is $\frac{\partial}{\partial q_j}(\mathbf{E}\{(p_i(q_i, q_j) - c - \gamma_i) \cdot q_i\})$, where $p_i(.)$ is the inverse demand curve for retailer i.

right pricing incentives to the retailers from the ex ante viewpoint of the industry. However, the expected profit for the manufacturer does not attain the first best level as the retailers do not face the right marginal cost from the industry's perspective when the retail costs are different from the expected ones: the wholesale price is based on the expected opportunity cost $\frac{1}{2}\beta(\frac{1}{1-\beta} - c - \gamma^e)$, not on the real opportunity cost $\frac{1}{2}\beta(\frac{1}{1-\beta} - c - \gamma_i)$.

The franchise fee serves to capture the entire resulting expected surplus at the retail level (leaving the retailers with an expected operating profit just enough to cover the fixed cost F). As the retailers are risk neutral, there are no specific insurance constraints and the producer is not hindered in choosing a franchise fee that captures all profits in expectation.

5.3.2 Resale price maintenance

With resale price maintenance, one of the sales conditions is that the product must be resold at the price that the producer has indicated. Denote this price by $\tilde{p}$. Let the wholesale tariff again be $T(q) = A + p_w \cdot q$, where A is the franchise fee, p_w the unit wholesale price and q the quantity sold to the retailer. Now, the manufacturer solves

$$\begin{aligned} &\max_{A,p_w,\tilde{p}} \mathbf{E}\{\pi_{prod}(\tilde{p}, p_w)\} + 2 \cdot A \\ &\text{s.t. } A \leq \mathbf{E}\{\pi_i(\tilde{p}, p_w)\} \end{aligned}$$

where $\pi_{prod}(\tilde{p}, p_w)$ is the producer's operating profit when it sets retail price $\tilde{p}$ and charges marginal wholesale price p_w,

$$\pi_{prod}(\tilde{p}, p_w) = (p_w - c) \cdot \sum_{i=1,2} d_i(\tilde{p}, \tilde{p}),$$

and $\pi_i(\tilde{p}, p_w)$ is the corresponding retailer's profit (excluding the franchise fee):

$$\pi_i(\tilde{p}, p_w) = (\tilde{p} - p_w - \gamma_i) \cdot d_i(\tilde{p}, \tilde{p}) - F$$

The opportunity for the manufacturer to use a franchise fee turns out to be immaterial in this case. For a given choice of retail price $\tilde{p}$, it can just choose p_w large enough to take away, in expectation, all surplus at the retail level. (A and p_w are indeterminate, as long as retailers obtain a non-negative profit in expectation). Hence, for the manufacturer, the problem comes down to choosing $\tilde{p}$ so as to maximize industry profit

$$\max_{\tilde{p}} \mathbf{E}\{\sum_{i=1,2}(\tilde{p}-c-\gamma_i)\cdot d_i(\tilde{p},\tilde{p})\}-2F$$

We obtain the following lemma.

Lemma 5.3 *(Risk neutral retailers).* Under resale price maintenance, the manufacturer optimally imposes a retail price equal to the expected price level of the first best: $p^{RPM}=\mathbf{E}\{p^I\}$.

Proof. See appendix.

By comparing the profit levels of the manufacturer under both regimes, the following proposition follows:

Proposition 5.2 *When retailers are risk neutral, the manufacturer adopts a policy of free retail competition and two-part tariffs.*

Proof. See appendix.

Close inspection of the profit functions reveals that the profit obtained under resale price maintenance ($\mathbf{E}\{\pi^{RPM}\}$) falls short of the profit obtained under free retail competition using two-part tariffs by an amount proportional to $(\bar{\gamma}-\underline{\gamma})^2$. This reflects the point that under free retail competition local cost information is used, whereas under resale price maintenance, the producer chooses a retail price ex ante and that, therefore, local information in the hands of the retailers remains unused. Naturally, the performance of resale price maintenance becomes worse as the variability of the cost parameters increases. In expectation, the retail prices under resale price maintenance and free retail competition are the same, but it is the use of local information that makes free retail competition better from the viewpoint of the manufacturer.

5.3.3 Welfare assessment

As long as the fixed costs of investment (F) are not too large, it is welfare optimal to have two retail outlets: consumers value retail variety. As for the welfare comparison between the regime of free retail competition and resale price maintenance, we can first note that in terms of the expected retail price, there is no difference between the two (as mentioned above). However, it is only under free retail competition that the quantity sold (i.e. consumption) varies: it goes up with lower retail costs γ and goes down with higher retail costs. Not only is this better for the industry (free retail competition gives a higher pay-off than resale price maintenance does), also

consumers prefer this variability. Let us have a precise look at the consumers' surplus. Total consumers' surplus is just the mathematical sum of the surpluses derived from consuming products i and j sold by the two retailers[12]: $CS = CS_i + CS_j$. The surplus derived from consuming product i is given by

$$CS_i(p_i, p_j) = \int_{p_i}^{\infty} d_i(x_i, p_j)dx_i = \frac{1}{2}(1 - p_i + \beta \cdot p_j)^2 = \frac{1}{2}q_i^2,$$

a familiar expression for the case of linear demand. Hence, in expectation, consumer surplus is

$$\mathbf{E}\{CS\} = 2 \cdot \mathbf{E}\{\frac{1}{2}q_i^2\} = (\mathbf{E}\{q_i\})^2 + V\{q_i\}$$

In short, consumer surplus increases with the average level of consumption $\mathbf{E}\{q_i\}$ and with the variance of consumption (the consumer surplus is convex in consumption). The conclusion is that with risk neutral retailers, total welfare is higher under free retail competition than under resale price maintenance. The private incentives of the manufacturer in its choice of distribution system are, therefore, in line with the social optimum[13]. We summarize the result of this section in the following proposition.

Proposition 5.3 *When retailers are risk neutral, the manufacturer adopts a policy (free retail competition) which also results in the highest welfare.*

5.4 The manufacturer's choice: the case of risk averse retailers

5.4.1 Two-part tariffs

Let us again suppose that the producer aims at having two retailers carrying its product (this assumption will be checked later in the appendix). When the retailers are infinitely risk averse they are only willing to accept the contract offered by the manufacturer when it gives them a non-negative profit also in the worst case scenario. Under idiosyncratic cost uncertainty, the worst case scenario for retailers is that their

[12] In general, this need not be true, but here it is because the demand functions exhibit no income effect (Tirole, 1988, p.9). A more general condition would be that the cross-partial derivatives of the demand functions are equal, a condition that is also met.

[13] Of course, the actual price level is still to high from the viewpoint of welfare maximization.

competitors turn out to be very agressive (their cost efficiency is strong, $\underline{\gamma}$), whereas they themselves are very inefficient (their cost parameter is $\bar{\gamma}$). Let us denote by $\underline{\pi}_i(p_w)$, the lowest profit retailer i can possibly get. It is obtained from expression (5.2) by inserting $\gamma_i = \bar{\gamma}, \gamma_j = \underline{\gamma}$.

In order to obtain a maximum profit, while inducing the retailers to participate, the manufacturer chooses franchise fee A and wholesale price p_w so as to solve the following program:

$$\begin{aligned} &\max_{A,p_w} \mathbf{E}\{\pi_{prod}(p_w)\} + 2 \cdot A \\ &\text{s.t. } A \leq \underline{\pi}_i(p_w) \end{aligned}$$

The maximum franchise fee that the manufacturer can charge is such that the retailers are left without profit, in the worst case scenario. This maximal fee is also the optimal one: it is not optimal for the producer to leave profit in the hands of the retailers and it cannot ask for more. Hence, the maximization problem reduces to

$$\max_{p_w} \mathbf{E}\{\pi_{prod}(p_w)\} + 2 \cdot \underline{\pi}_i(p_w)$$

We state the following lemma:

Lemma 5.4 *(Infinitely risk averse retailers)* Under free retail competition, the optimal tariff induces an expected retail price that is too high from the industry's point of view: $\mathbf{E}\{p_i\} > \mathbf{E}\{p_i^I\}$. Further, the manufacturer obtains a profit that is smaller than the one obtained when retailers are risk neutral.

Proof. See appendix.

The reason is simple. Because the retailers are risk averse, the manufacturer has to be considerate of the risk that they incur when a franchise fee is imposed on them. This impairs the manufacturer to impose the powerful tariff structure found in the preceding section that ensured optimal pricing incentives (in expectation) and allowed the manufacturer to capture the entire industry surplus. Under this tariff scheme, the manufacturer effectively 'sold out' the business to the retailers: the manufacturer obtained the entire expected industry surplus by an upfront payment (the franchise fee) and let the retailers take all the risk. With risk neutral resellers, this is no problem, with risk averse retailers it is.

5.4.2 Resale price maintenance

What outcome can be attained by the manufacturer when it applies resale price maintenance? Denote by $\tilde{p}$ the price that the manufacturer imposes. Let the wholesale tariff again be $T(q) = A + p_w \cdot q$. Now, the manufacturer solves

$$\max_{A,p_w,\tilde{p}} \mathbf{E}\{\pi_{prod}(\tilde{p}, p_w)\} + 2 \cdot A$$
$$\text{s.t. } A \leq \underline{\pi}_i(\tilde{p}, p_w)$$

where $\underline{\pi}_i(\tilde{p}, p_w)$ is the worst possible outcome that a retailer can have under the current regime: it is $\pi_i(\tilde{p}, p_w) = (\tilde{p} - p_w - \bar{\gamma}) \cdot d_i(\tilde{p}, \tilde{p}) - F$. The opportunity for the manufacturer to use a franchise fee turns out to be immaterial in this case as well. It may just as well choose the wholesale price such that the retailer is left with zero profit in the worst state of nature[14]. Let us take $p_w = \tilde{p} - \bar{\gamma} - F/d_i(\tilde{p}, \tilde{p})$. Now, for the manufacturer, the problem comes down to choosing $\tilde{p}$ so as to maximize profit

$$\max_{\tilde{p}} \; 2(\tilde{p} - \bar{\gamma} - \frac{F}{d_i(\tilde{p}, \tilde{p})} - c) \cdot d_i(\tilde{p}, \tilde{p})$$

We obtain the following lemma.

Lemma 5.5 *(Infinitely risk averse retailers)* Under resale price maintenance, the manufacturer imposes a retail price equal to $p^{RPM} = \frac{1}{2(1-\beta)} \cdot (1 - (1 - \beta)(c + \bar{\gamma})) > \mathbf{E}\{p_i^I\}$. The manufacturer obtains a profit that is smaller than the one obtained when retailers are risk neutral.

Proof. See appendix.

A comparison with the profit obtained under risk neutrality reveals that the manufacturer effectively focuses on the worst possible cost situation $\bar{\gamma}$, rather than on the expected cost situation γ^e.

The following proposition states the main result of the current chapter.

Proposition 5.4 *When cost uncertainty is idiosyncratic but moderate (the spread $\bar{\gamma} - \underline{\gamma}$ is smaller than a cut-off level $\Delta\gamma^{\pi}$) and retailers are infinitely risk averse, the manufacturer prefers resale price maintenance to free retail competition.*

[14] Things are different when there is also demand uncertainty at play. In that case, the franchise fee is optimally zero. After all, the manufacturer can then choose the highest possible wholesale price p_w thats puts the retailer's net profit equal to zero in the worst case scenario ($0 = \underline{\pi}_i(\tilde{p}, p_w)$), which has the advantage of also increasing the manufacturer's profit in other states of nature (states in which demand turns out to be high).

Proof. See appendix.

Two effects are at play. The fact that retailers are risk averse makes it difficult for the manufacturer to impose a franchise fee on them, a fee which would counter the double marginalisation problem. The problem is that, with a substantial fee, the retailers may not break even when the competitor turns out to be very efficient. Resale price maintenance does not have this insurance problem, because it automatically neutralizes the adverse effect of a competitor being more efficient: every retailer is required to charge the same price whatever the realisation of cost. However, the larger the cost variability (the larger the difference $\bar{\gamma} - \underline{\gamma}$), the more the inflexibility of resale price maintenance becomes a weak point. Local information is not used under resale price maintenance, which implies a poor exploitation of market power.

It turns out that for small cost variability, the insurance argument dominates (rendering resale price maintenance the most profitable instrument) and that for large cost variability, the flexibility argument dominates (making free retail competition the more profitable instrument). The reason for resale price maintenance to be superior for small cost variability is not trivial and may be traced back to the following. When there is no uncertainty at all, free retail competition and resale price maintenance yield exactly the same profit. In the case of free retail competition, the introduction of a little bit of uncertainty (a small difference $\bar{\gamma} - \underline{\gamma}$) necessitates a reduction in the franchise fee for both retailers: the price schedule becomes less powerful for both retailers. Resale price maintenance, on the other hand, may imply a retail price that is somewhat too large for one retailer and too low for the other, but the overall performance is to be preferred to the sure loss in franchise fees under direct competition. It is only when the cost variability is large, that the inflexibility of resale price maintenance starts to count.

5.4.3 Welfare assessment

As long as the fixed costs of investment, F, are not too large, it is welfare optimal to have two retail outlets[15]. Further, we can derive that, in expectation, the (fixed)

[15]Given that the retailers are risk averse, one cannot simply add up consumer surplus and the profits of the vertical structure to characterise total welfare. The welfare optimum can, however, be thought of as the situation which would be chosen by a social planner. In the case of risk averse retailers, the social planner would choose to maximize the sum of consumer surplus and profits, while totally insuring the retailers through redistribution. In this sense, as long as the fixed costs of investment are not too large, it is welfare optimal to have two retail outlets.

retail price under resale price maintenance is lower than that of direct competition: in expectation it does a better job in tackling the double marginalisation problem than direct competition. This speaks in favour of resale price maintenance, both from the viewpoint of the industry (see Proposition 3) and that of the consumers. However, again, it is only under direct competition that the quantity sold (i.e. consumption) can vary: it goes up with lower retail costs γ and goes down with higher retail costs. This is better both for the industry and for consumers. It follows that as long as retail cost variability is small, total welfare is larger under resale price maintenance than under direct competition. Let us call the level of the cost uncertainty under which the two regimes lead to equal welfare $\Delta\gamma^w$. We obtain the following proposition:

Proposition 5.5 *When retailers are infinitely risk averse, there are cases in which the manufacturer's preference for resale price maintenace is welfare optimal (for small cost uncertainty, $\bar{\gamma} - \underline{\gamma} < \Delta\gamma^w$) and when it is not (for intermediate levels, $\Delta\gamma^w < \bar{\gamma} - \underline{\gamma} < \Delta\gamma^\pi$). Whenever the manufacturer adopts a policy of free retail competition, that policy also results in the highest welfare.*

Proof. See appendix.

It follows that as long as retail cost variability is small enough, $\bar{\gamma} - \underline{\gamma} < \Delta\gamma^w$, total welfare is larger under resale price maintenance than under direct competition. In that case, the private incentives of the manufacturer are in line with the social optimum. The result that for intermediate levels the producer chooses resale price maintenance whereas this is not welfare optimal, has to do with the fact that the producer does not internalize the effects of its policy on consumers and retailers (who both value variability in prices).

5.5 Discussion

The main conclusion of this chapter is that when retailers face idiosyncratic cost uncertainty, it is conceivable that the use of resale price maintenance is a way to increase the manufacturer's profits. Necessary conditions under which this is the case are that 1) the retailers possess market power 2) the retailers are sufficiently risk averse and 3) the cost variability is not too large. As for the first element: the presence of market power necessitates, in the case of retail competition, the use of a powerful two-part tariff (a positive franchise fee and a low wholesale price) to counter the double

marginalisation problem due to retailer differentiation. When retailers are risk averse (element 2), it is difficult to impose a fee on them and to effectively counter this problem. Therefore, it is resale price maintenance that serves best to exploit market power and to insure the retailers, as long as the loss of flexibility under resale price maintenance is not too important (element 3). The above reasoning also applies even more so, when retailers need to invest before they can start their enterprise. After all, from the viewpoint of the retailers, a franchise fee and a fixed cost of investment are not different: both must be regained in the market place[16].

In fact, the argument made in this chapter bears an analogy to a different case that has been considered by Rey and Tirole (1986), that of market-wide uncertainty about demand conditions. Whereas RPM and free retail competition under two-part tariffs turn out to be equally effective when (risk averse) retailers are non-differentiated, RPM outperforms free retail competition when they are differentiated. The reason is, again, that retailer differentiation requires a powerful two-part tariff in the case of competition to counter the double marginalisation problem. The necessary franchise fee exposes the retailers to an additional risk, for which they must be compensated. RPM, in their model specification, does not entail this risk, making it more easily employable for the manufacturer[17].

A natural extension of the current model would be to consider multiproduct retailers (cf. Rey and Vergé (1999)) in an uncertain context. This line of research is particularly interesting as it is sometimes alleged that when a product is used as a *loss leader* (meaning that the product is priced even below cost in order to attract consumers into the retail premises), the size of the distribution network is put under pressure as the number of retailers willing to carry the product concerned is reduced. Extrapolation of the results in this chapter would suggest that as resale price maintenance protects retailers from the risk that other retailers use the product in question

[16] In a parallel paper, Verouden (1997) 'Resale price maintenance and outlet density' mimeo CentER, Tilburg University, the case of inelastic demands, spatial competition (localized market power) and positive fixed costs of investment is studied. In that setting, resale price maintenance is found to be optimal when, for a given cost differential, the worst case scenario is an 'exception' rather than the rule. In that case, the loss arising from the inflexibility of resale price information is fairly limited. The reasoning is, therefore, similar to the reasoning in the case of small cost variability considered here.

[17] In the setting of Rey and Tirole (1986), RPM allows the manufacturer to set the retail margin to zero, which is clearly optimal from a retailer insurance perspective. The zero margin is specific to the assumption that the fixed costs of setting up a retail outlet are zero. When there are fixed costs however, the retail margin cannot be set to zero, unless the manufacturer can use negative franchise fees (e.g. subsidies).

as a loss leader, it may induce retailers to carry the product in their shops and enable a good exploitation of market power. Indeed, uncertainty in this case is more about whether or not price cutting will occur than about the true cost of retailing. This feature may go into the direction of satisfying condition 3) mentioned above. It is recognized by the author that more specific research is necessary to see whether this inferrence is indeed valid.

5.6 Appendices

Derivation of consumer demand when there is no retail variety

The consumers value retail variety. If the product is sold by two retailers, market demand d_i for retailer i $(i = 1, 2)$ is modelled as

$$d_i(p_i, p_j) = 1 - p_i + \beta \cdot p_j \qquad (j \neq i, 0 \leq \beta < 1) \tag{5.5}$$

What is the proper demand specification when there is only one retail outlet selling the manufacturer's product? Suppose that there is a representative utility function $U(.)$ underlying the demand specification (5.5). Let us take it quasi-linear in money, m:

$$U(q_1, q_2) = m + u(q_1, q_2)$$

Under a budget constraint $m + p_1 \cdot q_1 + p_2 \cdot q_2 \leq \overline{y}$, where $\overline{y}$ is disposable income, utility maximization comes down to

$$\max_{q_1, q_2} u(q_1, q_2) - p_1 \cdot q_1 - p_2 \cdot q_2$$

This yields the following first order conditions (assuming that the expenses made on the product are small in comparison with budget $\overline{y}$):

$$\frac{\partial u}{\partial q_i} = p_i \qquad (i = 1, 2) \tag{5.6}$$

Rewriting the system of demand functions towards a system of inverse demand functions:

$$p_i = \frac{1}{1 - \beta^2}(1 + \beta - q_i - \beta \cdot q_j) \qquad (i = 1, 2; j \neq i) \tag{5.7}$$

Combining (5.6) and (5.7) gives us

$$u(q_1, q_2) = \frac{1}{1 - \beta^2}((1 + \beta)q_1 - \frac{1}{2}q_1^2 - \beta \cdot q_1 \cdot q_1 + (1 + \beta)q_2 - \frac{1}{2}q_2^2)$$

If good 2 is not present, the consumers maximize

$$\max_{q_1} u(q_1, 0) - p_1 \cdot q_1$$

This gives demand specification

$$d_i(p_i) = 1 + \beta - (1 - \beta^2) \cdot p_i \qquad (i = 1, 2)$$

Proofs of Proposition 5.1 and Lemma 5.1

Suppose the structure opens one outlet. For given cost parameters, it will maximize

$$\pi^I_{one}(p_i) = (p_i - c - \gamma_i) \cdot d_i(p_i) - F$$

The retail price that maximizes profit for the integrated firm with one outlet, p^I_{one}, is given by

$$p^I_{one} = \frac{1}{2}(c + \gamma_i) + \frac{1}{2(1-\beta)}$$

with profits

$$\pi^I_{one} = \frac{(1+\beta)(1-(1-\beta)(c+\gamma_i))^2}{4(1-\beta)} - F$$

Ex ante, the expected level of profits is $\mathbf{E}\{\pi^I_{one}\}$,where the expectation is taken over $\gamma_i \sim U(\underline{\gamma}, \bar{\gamma})$. Using $\mathbf{E}\{x^2\} = (\mathbf{E}\{x\})^2 + \mathbf{V}\{x\}$, we obtain

$$\begin{aligned}
\mathbf{E}\{\pi^I_{one}\} &= \frac{1+\beta}{4(1-\beta)} \cdot \mathbf{E}\{(1-(1-\beta)(c+\gamma_i))^2\} - F \\
&= \frac{1+\beta}{4(1-\beta)} \cdot [(1-(1-\beta)(c+\gamma^e))^2 + \mathbf{V}\{(1-\beta)(c+\gamma_i)\}] - F \\
&= \frac{1+\beta}{4(1-\beta)} \cdot (1-(1-\beta)(c+\gamma^e))^2 + \frac{1-\beta^2}{48}(\bar{\gamma} - \underline{\gamma})^2 - F
\end{aligned}$$

where $\gamma^e = \mathbf{E}\{\gamma_i\}$.

Similarly, if the structure opens two outlets it will maximize

$$\pi^I_{two}(p_i, p_j) = \sum_{i=1, j\neq i}^{2} (p_i - c - \gamma_i) \cdot d_i(p_i, p_j) - 2F$$

which gives

$$p^I_i = \frac{1}{2}(c + \gamma_i) + \frac{1}{2(1-\beta)} \qquad (i = 1, 2)$$

Profits ex post for the integrated firm are

$$\pi^I_{two} = \frac{1}{4(1-\beta)} \cdot \sum_{i=1}^{2} (1-(1-\beta)(c+\gamma_i))^2 + \frac{1}{4}\beta(\bar{\gamma} - \underline{\gamma})^2 - 2F$$

In expectation, the level of profits is

$$\mathbf{E}\{\pi^I_{two}\} = \frac{1}{2(1-\beta)} \cdot (1-(1-\beta)(c+\gamma^e))^2 + \frac{1}{24}(\bar{\gamma}-\underline{\gamma})^2 - 2F$$

One can observe that for given γ^e, both $\mathbf{E}\{\pi^I_{two}\}$ and $\mathbf{E}\{\pi^I_{one}\}$ are increasing in the spread $(\bar{\gamma}-\underline{\gamma})$. In fact, the difference $\mathbf{E}\{\pi^I_{two}\} - \mathbf{E}\{\pi^I_{one}\}$ is minimal for $\bar{\gamma} = \underline{\gamma}$ (the case of no uncertainty), so we can concentrate on this case in order to obtain a sufficient condition for $\mathbf{E}\{\pi^I_{two}\} \geq \mathbf{E}\{\pi^I_{one}\}$. Working out the inequality $\pi^I_{two} \geq \pi^I_{one}$ for given γ $(=\gamma^e)$ gives the condition $F \leq \frac{1}{4} \cdot (1-(1-\beta)(c+\gamma^e))^2$.

Proof of Lemma 5.2

The manufacturer charges a franchise fee A equal to the retailer's expected profit $\mathbf{E}\{\pi_i(p_w)\}$ and chooses wholesale price p_w so as to solve the sum of (expected) operating profits and franchise fees:

$$\max_{p_w} \mathbf{E}\{\pi_{prod}(p_w)\} + 2 \cdot \mathbf{E}\{\pi_i(p_w)\}$$

When two firms enter the market, the Nash equilibrium prices and demand are

$$\begin{aligned} p_i(p_w) &= \frac{1}{2-\frac{1}{2}\beta^2}(1+p_w+\gamma_i+\frac{1}{2}\beta(1+p_w+\gamma_j)) \\ d_i(p_w) &= \frac{1}{2(1-\frac{1}{2}\beta)}(1-(1-\beta)(p_w+\gamma_i)+\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\gamma_j-\gamma_j)) \end{aligned}$$

The producer's operating profits are, in expectation,

$$\mathbf{E}\{\pi_{prod}(p_w)\} = (p_w - c) \cdot \frac{1}{(1-\frac{1}{2}\beta)}(1-(1-\beta)(p_w+\gamma^e))$$

The franchise fee is

$$\begin{aligned} \mathbf{E}\{\pi_i(p_w)\} &= \frac{1}{4(1-\frac{1}{2}\beta)^2} \cdot \mathbf{E}\{(1-(1-\beta)(p_w+\gamma_i)+\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\gamma_j-\gamma_i))^2\} - F \\ &= \frac{1}{4(1-\frac{1}{2}\beta)^2} \cdot [(\mathbf{E}\{(1-(1-\beta)(p_w+\gamma_i)+\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\gamma_j-\gamma_i))\})^2 + \\ &\quad \mathbf{V}\{(1-(1-\beta)(p_w+\gamma_i)+\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\gamma_j-\gamma_i))\}] - F \\ &= \frac{1}{4(1-\frac{1}{2}\beta)^2} \cdot [(1-(1-\beta)(p_w+\gamma^e))^2 + \frac{1-\frac{3}{4}\beta^2+\frac{1}{4}\beta^4}{(1+\frac{1}{2}\beta)^2}\frac{1}{12}(\overline{\gamma}-\underline{\gamma})^2] - F \end{aligned}$$

where it is used that $\mathbf{E}\{x^2\} = (\mathbf{E}\{x\})^2 + \mathbf{V}\{x\}$ and $\mathbf{V}\{\gamma_i\} = \frac{1}{12}(\overline{\gamma}-\underline{\gamma})^2$.

The first order condition of the maximization problem is

$$\frac{1}{(1-\frac{1}{2}\beta)}(1-(1-\beta)(p_w+\gamma^e)-(1-\beta)(p_w-c))- \\ \frac{1}{2(1-\frac{1}{2}\beta)^2}(2(1-(1-\beta)(p_w+\gamma^e)))(1-\beta)=0$$

which reduces to

$$p_w = c + \frac{1}{2}\beta(\frac{1}{1-\beta} - c - \gamma^e)$$

Total expected profits for the manufacturer are

$$\mathbf{E}\{\pi^A\} = \frac{1}{2(1-\beta)} \cdot (1-(1-\beta)(c+\gamma^e))^2 + \frac{1-\frac{3}{4}\beta^2+\frac{1}{4}\beta^4}{24(1+\frac{1}{2}\beta)^2}(\overline{\gamma}-\underline{\gamma})^2 - 2F$$

Note that the optimal wholesale price exceeds marginal cost c by an amount $\frac{1}{2}\beta(\frac{1}{1-\beta} - c - \gamma^e)$. This reflects the expected influence of selling one more unit to one retailer on the retail price (and the profits) of the other retailer. Using expression (5.7), the system of inverse demand functions corresponding to $d_i(p_i,p_j) = 1 - p_i + \beta \cdot p_j$, $(i = 1,2; j \neq i)$, we obtain that, evaluated at the level of the optimal wholesale price found above:

$$\begin{aligned}\frac{\partial}{\partial q_j}(\mathbf{E}\{(p_i(q_i,q_j) - c - \gamma_i) \cdot q_i\}) &= -\frac{\beta}{1-\beta^2}\mathbf{E}\{q_i\} \\ &= -\frac{\beta}{1-\beta^2}\frac{1}{2}(1-(1-\beta)(c+\gamma^e)) \\ &= -\frac{1}{2}\beta(\frac{1}{1-\beta} - c - \gamma^e)\end{aligned}$$

The optimal wholesale price indeed induces, in expectation, the first best retail price (the retail price that is optimal from the viewpoint of the industry):

$$\mathbf{E}\{p_i\} = \frac{1}{2(1-\beta)} \cdot (1+(1-\beta)(c+\gamma^e)) = \mathbf{E}\{p^I\}$$

Generically, it does not, however, induce a first best retail price for every realisation of the cost parameters:

$$\begin{aligned}p_i &= \frac{1}{2(1-\beta)} \cdot (1+(1-\beta)(c+\gamma_i)) + \frac{\frac{1}{2}\beta}{2-\frac{1}{2}\beta^2}(\frac{1}{2}\beta(\gamma_i-\gamma^e)+(\gamma_j-\gamma^e)) \\ &= p^I + \frac{\frac{1}{2}\beta}{2-\frac{1}{2}\beta^2}(\frac{1}{2}\beta(\gamma_i-\gamma^e)+(\gamma_j-\gamma^e)) \neq p^I\end{aligned}$$

Proof of Lemma 5.3

Under resale price maintenance, the manufacturer chooses a retail price $\tilde{p}$ so as to maximize expected profit

$$\max_{\tilde{p}} \mathbf{E}\{(\tilde{p}_i - c - \gamma_i) \cdot (1 - p_i + \beta p_j) + (\tilde{p}_j - c - \gamma_j) \cdot (1 - p_j + \beta p_i)\} - F$$

The first order conditions give

$$\tilde{p}_i = \tilde{p}_j = \frac{1}{2(1-\beta)} \cdot (1 + (1-\beta)(c + \gamma^e))$$

Expected profits for the producer:

$$\mathbf{E}\{\pi^{RPM}\} = \pi^{RPM} = \frac{1}{2(1-\beta)} \cdot (1 - (1-\beta)(c + \gamma^e))^2 - 2F$$

Proof of Proposition 5.2

By comparing the expected profit level of the manufacturer under the regime of free retail competition ($\mathbf{E}\{\pi^A\}$) with that under resale price maintenance (π^{RPM}), we obtain:

$$\mathbf{E}\{\pi^A\} - \pi^{RPM} = \frac{1 - \frac{3}{4}\beta^2 + \frac{1}{4}\beta^4}{24(1 + \frac{1}{2}\beta)^2}(\overline{\gamma} - \underline{\gamma})^2 > 0$$

Proof of Lemma 5.4

The maximum franchise fee that the manufacturer can charge when retailers are infinitely risk averse is $\underline{\pi}_i(p_w)$:

$$\underline{\pi}_i(p_w) = \frac{1}{4(1 - \frac{1}{2}\beta)^2}(1 - (1-\beta)(p_w + \overline{\gamma}) - \frac{\frac{1}{2}\beta}{1 + \frac{1}{2}\beta}(\overline{\gamma} - \underline{\gamma}))^2 - F$$

The producer's operating profits are, in expectation,

$$\mathbf{E}\{\pi_{prod}(p_w)\} = (p_w - c) \cdot \frac{1}{(1 - \frac{1}{2}\beta)}(1 - (1-\beta)(p_w + \gamma^e))$$

The maximization problem is

$$\max_{p_w} \mathbf{E}\{\pi_{prod}(p_w)\} + 2 \cdot \underline{\pi}_i(p_w)$$

First order condition:

$$\frac{1}{(1-\frac{1}{2}\beta)}(1-(1-\beta)(p_w+\gamma^e)-(1-\beta)(p_w-c))- \\ \frac{1}{2(1-\frac{1}{2}\beta)^2}(2(1-(1-\beta)(p_w+\overline{\gamma})-\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\overline{\gamma}-\underline{\gamma})))(1-\beta)=0$$

which gives the following optimal wholesale price:

$$p_w=(1-\frac{1}{2}\beta)\cdot c+\frac{1}{2}(1-\frac{1}{2}\beta)(\overline{\gamma}-\underline{\gamma})-\frac{\frac{1}{4}\beta^2}{1+\frac{1}{2}\beta}\overline{\gamma}-\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}\underline{\gamma}+\frac{\frac{1}{2}\beta}{1-\beta} \tag{5.8}$$

and the corresponding franchise fee

$$A=\underline{\pi}_i(p_w)=(1-(1-\beta)(c+\overline{\gamma})-\frac{\frac{1}{2}\beta^2-\frac{1}{2}\beta-1}{2(1+\frac{1}{2}\beta)}(\overline{\gamma}-\underline{\gamma}))^2-F \tag{5.9}$$

After tedious calculations and some rewriting, we obtain the resulting total expected profit for the manufacturer under the two-part tariff

$$\begin{aligned}\mathbf{E}\{\pi^A\} &= \frac{1}{2(1-\beta)}\cdot(1-(1-\beta)(c+\overline{\gamma}))^2-\frac{\frac{1}{2}\beta}{2(1+\frac{1}{2}\beta)}(1-(1-\beta)(c+\overline{\gamma}))(\overline{\gamma}-\underline{\gamma}) \\ &\quad +\frac{1+2\beta+\frac{5}{4}\beta^2-\frac{1}{4}\beta^3}{2(2(1+\frac{1}{2}\beta))^2}(\overline{\gamma}-\underline{\gamma})^2-2F\end{aligned}$$

Proof of Lemma 5.5

With infinitely risk averse retailers, the manufacturer chooses retail price $\tilde{p}$ so as to solve

$$\max_{\tilde{p}}\ 2(\tilde{p}-\bar{\gamma}-\frac{F}{d_i(\tilde{p},\tilde{p})}-c)\cdot d_i(\tilde{p},\tilde{p})$$

This correspronds to solving

$$\max_{\tilde{p}}\ 2(\tilde{p}-\bar{\gamma}-c)\cdot(1-(1-\beta)\tilde{p})-2F$$

which gives

$$\tilde{p}_i=\tilde{p}_j=\frac{1}{2(1-\beta)}\cdot(1+(1-\beta)(c+\bar{\gamma}))$$

The resulting profits for the manufacturer:

$$\mathbf{E}\{\pi^{RPM}\}=\pi^{RPM}=\frac{1}{2(1-\beta)}\cdot(1-(1-\beta)(c+\bar{\gamma}))^2-2F$$

Proof of Proposition 5.4

Comparing the (expected) profits under free retail competition (two-part tariffs) and under resale price maintenance gives

$$\begin{aligned}\mathbf{E}\{\pi^{RPM}\} - \mathbf{E}\{\pi^{A}\} &= \frac{\frac{1}{2}\beta}{2(1+\frac{1}{2}\beta)}(1-(1-\beta)(c+\overline{\gamma}))(\overline{\gamma}-\underline{\gamma}) - \\ &\quad \frac{1+2\beta+\frac{5}{4}\beta^2-\frac{1}{4}\beta^3}{2(2(1+\frac{1}{2}\beta))^2}(\overline{\gamma}-\underline{\gamma})^2 \\ &= \frac{\frac{1}{2}}{2(1+\frac{1}{2}\beta)}[\beta(1-(1-\beta)(c+\overline{\gamma})) - \\ &\quad \frac{1+2\beta+\frac{5}{4}\beta^2-\frac{1}{4}\beta^3}{4(1+\frac{1}{2}\beta)}(\overline{\gamma}-\underline{\gamma})](\overline{\gamma}-\underline{\gamma})\end{aligned}$$

which is positive if and only if

$$\beta(1-(1-\beta)(c+\overline{\gamma})) - \frac{1+2\beta+\frac{5}{4}\beta^2-\frac{1}{4}\beta^3}{4(1+\frac{1}{2}\beta)}(\overline{\gamma}-\underline{\gamma}) > 0$$

This is the case for a cost variability $(\overline{\gamma}-\underline{\gamma})$ that is not too large:

$$\overline{\gamma}-\underline{\gamma} < \frac{4\beta(1+\frac{1}{2}\beta)}{1+2\beta+\frac{5}{4}\beta^2-\frac{1}{4}\beta^3}\cdot(1-(1-\beta)(c+\overline{\gamma})) =: \Delta\gamma^{\pi}$$

Proof of Proposition 5.5

Let us first consider the consumers. In expectation, total consumer surplus is

$$\mathbf{E}\{CS\} = 2\cdot\mathbf{E}\{\frac{1}{2}q_i^2\} = (\mathbf{E}\{q_i\})^2 + \mathbf{V}\{q_i\}$$

Under resale price maintenance, retail price and demand $(q_i = \frac{1}{2}(1-(1-\beta)(c+\bar{\gamma})))$ are fixed, so that

$$\mathbf{E}\{CS^{RPM}\} = \frac{1}{4}(1-(1-\beta)(c+\bar{\gamma}))^2$$

Under downstream competition (with two-part tariffs), we have

$$q_i = \frac{1}{2(1-\frac{1}{2}\beta)}(1-(1-\beta)(p_w+\gamma_i)) + \frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\gamma_j-\gamma_i))$$

where p_w is given by expression (5.8). In expectation,

$$\mathbf{E}\{q_i\} = \frac{1}{2}(1-(1-\beta)(c+\overline{\gamma}) - \frac{\frac{1}{4}\beta(1-\beta)}{1+\frac{1}{2}\beta}(\overline{\gamma}-\underline{\gamma}))$$

Clearly, this expected quantity is lower than the quantity that results under resale price maintenance. As for the variability:

$$\begin{aligned}
\mathbf{V}\{q_i\} &= \frac{1}{2(1-\frac{1}{2}\beta)^2}\mathbf{V}\{-\gamma_i+\beta\gamma_i-\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}\gamma_i+\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}\gamma_j\} \\
&= \frac{1}{2(1-\frac{1}{2}\beta)^2}(\frac{(1-\frac{1}{2}\beta^2)^2}{(1+\frac{1}{2}\beta)^2}+\frac{\frac{1}{4}\beta^2}{(1+\frac{1}{2}\beta)^2})\mathbf{V}\{\gamma_i\} \\
&= \frac{1-\frac{3}{4}\beta^2+\frac{1}{4}\beta^4}{24(1-\frac{1}{4}\beta^2)^2}(\overline{\gamma}-\underline{\gamma})^2
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{E}\{CS^A\} &= \frac{1}{4}(1-(1-\beta)(c+\overline{\gamma})-\frac{\frac{1}{4}\beta(1-\beta)}{1+\frac{1}{2}\beta}(\overline{\gamma}-\underline{\gamma}))^2+\frac{1-\frac{3}{4}\beta^2+\frac{1}{4}\beta^4}{24(1-\frac{1}{4}\beta^2)^2}(\overline{\gamma}-\underline{\gamma})^2 \\
&= \frac{1}{4}(1-(1-\beta)(c+\overline{\gamma}))^2-\frac{\frac{1}{4}\beta(1-\beta)}{1+\frac{1}{2}\beta}(1-(1-\beta)(c+\overline{\gamma}))(\overline{\gamma}-\underline{\gamma}))+ \\
&\quad \frac{1-\frac{9}{4}\beta^3+\frac{17}{4}\beta^4-\frac{9}{8}\beta^5+\frac{3}{16}\beta^6}{24(1-\frac{1}{4}\beta^2)^2}(\overline{\gamma}-\underline{\gamma})^2
\end{aligned}$$

Let us now consider the retailers. Unlike the case of risk neutrality, retailers do obtain a positive expected rent when they are risk averse. For RPM,

$$\begin{aligned}
\pi_{_i}(\tilde{p},p_w) &= (\tilde{p}-p_w-\gamma_i)\cdot(1-(1-\beta)\tilde{p})-F \\
&= (\frac{F}{1-(1-\beta)\tilde{p}}+\overline{\gamma}-\gamma_i)\cdot(1-(1-\beta)\tilde{p})-F \\
&= \frac{1}{2}(1-(1+(1-\beta)(c+\bar{\gamma}))\cdot(\overline{\gamma}-\gamma_i)
\end{aligned}$$

and so

$$\mathbf{E}\{\pi_{_i}(\tilde{p},p_w)\} = \frac{1}{4}(1-(1+(1-\beta)(c+\bar{\gamma}))\cdot(\overline{\gamma}-\underline{\gamma})$$

For the case of downstream competition:

$$\pi_i(p_w) = \frac{1}{4(1-\frac{1}{2}\beta)^2}(1-(1-\beta)(p_w+\gamma_i)+\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\gamma_j-\gamma_i))^2-A-F$$

In expectation,

$$\begin{aligned}
\mathbf{E}\{\pi_i(p_w)\} &= \frac{1}{4(1-\frac{1}{2}\beta)^2}(1-(1-\beta)(p_w+\gamma^e))^2 + \\
&\quad \frac{1}{4(1-\frac{1}{2}\beta)^2}\cdot\mathbf{V}\{1-(1-\beta)(p_w+\gamma_i)+\frac{\frac{1}{2}\beta}{1+\frac{1}{2}\beta}(\gamma_j-\gamma_i)\} - A - F \\
&= \frac{1}{4}(1-(1-\beta)(c+\overline{\gamma})-\frac{\frac{1}{2}\beta}{2(1+\frac{1}{2}\beta)}(\overline{\gamma}-\underline{\gamma}))^2 + \\
&\quad \frac{1-\frac{3}{4}\beta^2+\frac{1}{4}\beta^4}{48(1-\frac{1}{4}\beta^2)^2}(\overline{\gamma}-\underline{\gamma})^2 - A - F \\
&= -\frac{\frac{1}{4}\beta(1-\beta)}{2(1+\frac{1}{2}\beta)}(1-(1-\beta)(c+\overline{\gamma}))\cdot(\overline{\gamma}-\underline{\gamma}) + \\
&\quad \frac{1-\frac{9}{4}\beta^3+\frac{43}{16}\beta^4-\frac{9}{8}\beta^5+\frac{3}{16}\beta^6}{48(1-\frac{1}{4}\beta^2)^2}(\overline{\gamma}-\underline{\gamma})^2
\end{aligned}$$

with expression (5.9) substituted for A.

Combining the results for the producer surplus (Proposition 5.4), the consumer surplus and the surplus of the retailers, we obtain

$$\begin{aligned}
\mathbf{E}\{W^{RPM}\} - \mathbf{E}\{W^A\} &= \frac{\frac{1}{4}\beta(1-\beta)}{1+\frac{1}{2}\beta}(1-(1-\beta)(c+\overline{\gamma}))\cdot(\overline{\gamma}-\underline{\gamma}) \\
&\quad -\frac{\frac{8}{3}-4\beta+4\beta^2-2\beta^3+\frac{17}{4}\beta^4-\frac{3}{4}\beta^5+\frac{1}{4}\beta^6}{32(1-\frac{1}{4}\beta^2)^2}\cdot(\overline{\gamma}-\underline{\gamma})^2
\end{aligned}$$

which is positive if and only if

$$\overline{\gamma}-\underline{\gamma} < \frac{8\beta(1-\beta)(1+\frac{1}{2}\beta)(1-\frac{1}{2}\beta^2)}{\frac{8}{3}-4\beta+4\beta^2-2\beta^3+\frac{17}{4}\beta^4-\frac{3}{4}\beta^5+\frac{1}{4}\beta^6}\cdot(1-(1-\beta)(c+\overline{\gamma})) =: \Delta\gamma^w$$

It is easily verified that $\Delta\gamma^w > 0$.

5.7 References

Rey, P. and Tirole, J. (1986) 'The logic of vertical restraints' American Economic Review 76: 921-936.

Rey, P. and Tirole, J. (1997) 'A primer on foreclosure' mimeo IDEI, Université de Toulouse. Forthcoming in Brookings Papers on Economic Activity.

Rey, P. and Vergé, T. (1999) 'Resale price maintenance and horizontal collusion' mimeo IDEI, Université de Toulouse.

Segal, I. (1999) 'Contracting with externalities' Quarterly Journal of Economics CXIVV(2): 337-388.

Tirole, J. (1988) 'The theory of industrial organization' MIT Press, Cambridge MA, USA.

Chapter 6

Cartel formation under incomplete information: on the requirement of collusion-proofness

.

6.1 Introduction

This chapter is about cartel formation[1]. It is generally thought that the likelihood of firms forming a cartel is greater in concentrated industries than in industries with many firms. Not only because it is, so the argument goes, easier to monitor a cartel agreement in the relatively surveyable environment of a tight oligopoly (cartel enforcement argument) but also because it may be easier or more attractive for fewer firms to come to terms about the conditions applying to the cartel (cartel formation argument).

One element that can be a source of difficulty in the formation of cartels is the problem of *incomplete information* with respect to the cost levels of the participating firms. This information asymmetry may pose a problem at the stage where the cartel must determine the conditions of the cartel agreement (e.g. production quotas, fixed market shares) for the participants. Obviously, the conditions of the cartel agreement also bear on the decision to join the cartel in the first place. For example, a firm that is relatively efficient will typically only be satisfied with a production quota that

[1]This chapter is jointly written with Jérôme Pouyet from the University of Toulouse.

somehow steps up to this fact (in other words, the quota must be relatively large), otherwise it may just prefer to compete on the market. This, however, should induce firms which are less efficient to overstate their efficiency in order to obtain a higher share of the cartel output. But when every firm is saying to be efficient (or saying to have become more efficient since the latest negotiations) and claiming large quota, this will reduce the attractiveness of the cartel for firms which are effectively among the most efficient. An illustration of the ensuing difficulties is given by Eckbo (1976), who found that in a sample of international cartels which were temporarily successful but then broke down, almost half of them were finished due to internal squabbling over how to share the profits.

The extent to which cartel agreements can overcome the conflicting requirements mentioned above has been the subject of extensive research, cf. Roberts (1985), Kihlstrom and Vives (1989,1992) and Cramton and Palfrey (1990). In order to characterise the outcomes that a cartel can achieve in situations of incomplete information, these authors have approached the issue of cartel formation using a standard mechanism design approach: in a (Bayesian) Cournot industry, there is a 'cartel manager' who proposes a cartel arrangement and determines the optimal quotas depending on the costs each firm announces to have[2]. Given this scheme, firms decide whether or not to join and, if they do, they announce their costs. A proposed cartel agreement is called 'efficient' when only the firm(s) with the lowest cost produce(s). In order to form such a cartel, the cartel manager must, according to the well-known Revelation Principle, propose a scheme (possibly involving side payments) that ensures individual participation (Bayesian individual rationality, BIR) and induces the firms to individually reveal their cost information (Bayesian incentive compatibility, BIC).

In a setting where the number of possible efficiency types is limited to two (firms are either efficient or inefficient), Kihlstrom and Vives (1989, 1992) have shown that the formation of an efficient cartel *is* possible, both in the case of a duopoly and in the borderline case of an industry comprising infinitely many firms (modelled as a continuum of firms). The reason of this latter, rather surprising, result is that in an

[2]In studying the problems related to cartel formation, the literature abstracts away from the cartel enforcement problem (and vice versa). Cartel arrangements are assumed to be enforceable, even though the secrecy of the arrangement implies that there is, for example, no public authority available to enforce it. The assumption of enforceability is a short-cut to capture in a static context the reputation of the cartel manager and the firms participating in the cartel which guarantee the self-enforceability of the arrangement in a dynamic context (e.g. by means of trigger strategies; Folk Theorem).

atomistic industry there is no uncertainty about the type of firm that should produce nor about the fraction of efficient firms being present, so that it does not turn out to be so difficult to reconcile all individual participation and incentive requirements. This result holds for all meaningful probability distributions on the two cost types.

Cramton and Palfrey (1990), however, have shown that if the private information of the firms can take a continuum of values (ranging from being very efficient to being very inefficient), the first best outcome can only be implemented if there are not too many firms in the industry. If there are many firms, they show that it becomes increasingly difficult to reconcile all the individual incentive and participation requirements outlined above. In this sense, their main result is a confirmation of the general idea that it is easier to form a cartel in an industry with a few firms than in one with many. Nonetheless, the extent to which their result is driven by assuming the uniform probability distribution (assigning equal probability to all possible cost types, even the most extreme values) is not clear. Furthermore, the assumption that the private information of the firms can take a continuum of values leads to the rather peculiar result that when less than unanimous consent is needed to ratify the cartel agreement (only a proportion $\alpha < 1$ of firms must agree with it), the impossibility result is lost. In other words, with less than unanimous consent, an efficient cartel can be implementable in industries with many participants[3].

The current literature, therefore, does seem to provide some justification for the generally held belief that cartels are most difficult to form in industries with many firms, but it fails to do so in several constellations, most notably in the context of a finite number of cost types with general probability distributions. In this chapter we will again consider the issue of cartel formation in this latter context, but from a different angle: in order to characterise the possible outcomes that a cartel can achieve, we propose to explore the additional requirement of *collusion-proofness*. The above mentioned models of the cartel manager (the principal) trying to obtain the efficient cartel outcome by inducing truthful cost revelation by the firms (the agents)

[3]Cramton and Palfrey show that when less than unanimous consent is needed to ratify the cartel agreement, an efficient cartel is implementable, *especially* when the number of firms is very large. This is due to a special feature of the continuum type model, namely the fact that when the number of firms grows, the lowest cost firm that is supposed to produce the entire cartel output has an increasingly small measure (it is one of the many firms). As in the limit, it is only the lowest cost firm whose participation constraint cannot be met in the face of the incentive requirements with respect to the less efficient firms, this firm will not be able to stop the cartel from going ahead if a majority rule applies.

all use the standard assumption that every agent behaves non-cooperatively: no communication is possible between the agents, which is a standard assumption for the Revelation Principle. The aim of the current chapter is to see whether the obtained results continue to hold when communication between groups of firms *cannot* be excluded and, in particular, when groups of firms try to (secretly) *coordinate* their cost announcements in order to obtain a better result.

So as to investigate the role of this possibility, we will model this kind of collusion following Laffont and Martimort (1997, 2000): in a first stage, the cartel manager (the principal) proposes a cartel contract ('grand mechanism') (q, t), where q is the vector of quantities to be produced by the cartel members and t the vector of side transfers as functions of the cost messages[4]. If any firm refuses this contract, the firms will compete à la Cournot. In phase 2, a 'mediator' (the third party) proposes to a group of firms a side contract (ϕ, y), where ϕ is the vector of manipulated cost messages of the considered group to be sent to the principal and y the vector of internal side payments. Finally, the firms in the group decide whether or not to accept the third party's program. If not, they will non-cooperatively send messages to the principal.

In the framework of Laffont and Martimort (1997, 2000), where, by assumption, collusion can only happen by all agents together, the implementable contracts can be characterised using the *Collusion-Proofness Principle*. This principle states that the cartel manager can, without loss of generality, restrict the cartel contract which he proposes to be a mechanism such that no collusion takes place at the equilibrium. In our setting, where collusion may take place by coalitions of variable size, we will determine to what extent collusion-proofness is relevant.

The possibility of collusion by subcoalitions is shown to change the set of implementable rules, but *not* to change the principal result that efficient cartel formation is possible for any number of firms in the industry. Partly, this is due to the strong congruence of interest between the cartel manager and the individual cartel members: after all, the cartel manager is acting costlessly on behalf of the members, by maximizing their total expected profits. But there is more to it. Typically, the transfer schemes which are able to implement collusion-proof cartels appear to satisfy a property which we call *partial anonymity:* the cartel transfers to the inefficient firms

[4]Instead of an explicit contract, one may also think of an arrangement that is enforceable in a dynamic context by means of trigger strategies (see footnote 2). The transfers will, for simplicity, be considered as monetary transfers. In principle, however, they can stand for any kind of compensation, e.g. promising not to enter a different geographical market where the other firms are already active.

are taken not to depend on the number of inefficient firms present in the cartel. The following intuition is then at the heart of the result: consider an inefficient firm which is instructed by the third party to represent itself to the cartel manager as being efficient. This firm will then get more revenue out of production, but will at the same time exert a negative externality on all firms that are truly efficient, including the efficient firms within the subcoalition (as the cartel output must now be divided over a larger number of firms). If, furthermore, the cartel transfers are partially anonymous, then the effect of the contemplated action is neutral with respect to the other inefficient firms in the subcoalition. So we are left to compare the direct effect on the pay-off of the particular inefficient firm with the negative externality on the efficient firms in the subcoalition. From this it follows that as soon as the inefficient firm is made individually indifferent between lying and truthtelling, the group incentive constraints will be satisfied. This line of reasoning identifies a transfer scheme that enables the implementation of cartels which are not subject to internal manipulation, in industries of any size. Finally, we identify cases in which the extra requirement of collusion-proofness need not have an impact on the minimal level of transfers that is required to form these cartels.

6.2 The model

We consider the possibilities for cartel formation in an industry with n firms producing a homogeneous good. Let q^i denote the production level of firm $i \in N = \{1, ..., n\}$ and $Q = \sum_{i=1}^{n} q^i$ the total market production. As in Kihlstrom and Vives (1989, 1992) and Cramton and Palfrey (1990), market demand for the good is given by a linear inverse demand function $p(Q) = 1 - Q$. The firms in the industry can be of two types: either efficient (their unit cost of production is low, $\underline{c}$) or inefficient ($\overline{c}$). Let Δc denote the cost difference $\overline{c} - \underline{c}$. The levels of efficiency are private knowledge to the firms, they are not observable to other firms or to the cartel (embodied by the cartel manager). Nonetheless, it is common knowledgde that the probability of an individual firm being efficient is $\mu \in (0, 1)$ and that the types are independently and identically distributed (i.i.d.). The output levels of the individual firms are observable.

The cartel manager proposes the firms a cartel contract $\{q_i, t_i\}_{i \in N}$, where $q_i = q_i(\tilde{c}^N)$ is the quantity to be produced by firm i and $t_i = t_i(\tilde{c}^N)$ is the transfer that firm i has to give (if $t_i > 0$) or receives from the cartel manager (if $t_i < 0$) as functions of the cost reports $\tilde{c}^N$. The transfers are restricted to be budget balanced in each state

of nature.

The gain of firm i is given by

$$\pi^i(\tilde{c}^N) = [p(Q(\tilde{c}^N)) - c^i]q^i(\tilde{c}^N) - t^i(\tilde{c}^N).$$

The objective of the cartel manager is to form a cartel that is *efficient*, as defined as follows:

Definition 6.1 *An efficient cartel is a cartel that maximizes the industry profit* $\sum_{i\in N} \pi^i(\tilde{c}^N)$ *in each state of nature.*

A proposed cartel agreement is efficient when only the firm(s) with the lowest cost produce(s) and the total industry output equals the monopoly output for this cost level, being $q^m(c) = \arg\max_q\{(p(q) - c)q\}$. A symmetric and efficient cartel amounts to the following rule chosen by the cartel manager

$$q^i(\tilde{c}^N) = \begin{cases} \frac{1}{j}q^m(\underline{c}) & \text{if } \tilde{c}^i = \underline{c}, \\ 0 & \text{if } \tilde{c}^i = \overline{c} \text{ and } j \neq 0 \\ \frac{1}{n}q^m(\overline{c}) & \text{if } \tilde{c}^i = \overline{c} \text{ and } j = 0 \end{cases}$$

where j is the total number of firms that report to be efficient to the cartel manager[5]. Let $t(j)$ be the total transfer to be paid by the efficient firms (and consequently to be received by the inefficient firms). Then, we have

$$t^i(\tilde{c}^N) = \begin{cases} -\frac{1}{j}t(j) & \text{if } \tilde{c}^i = \underline{c}, \\ \frac{1}{n-j}t(j) & \text{if } \tilde{c}^i = \overline{c}. \end{cases}$$

Notice that when every firm reports to be efficient ($j = n$) or every firm reports to be inefficient ($j = 0$), no transfers will occur: $t(n) = t(0) = 0$. For the rest of the chapter, we will use the following conventions: all the transfers $t(j)$ are non-negative. This means that, unless the transfers are zero, the efficient firms will pay monetary transfers to the cartel manager, which are then redistributed to the inefficient firms. This is an innocuous assumption because we are primarily interested in efficient cartels in which only the most efficient firms produce and consequently derive gains from

[5]In a symmetric equilibrium, to characterise the state of nature it is enough to know the number of efficient (or inefficient) firms in the cartel.

production. In order that the inefficient firms are willing to take part in the cartel, they must normally be compensated for not producing.

6.3 The individual constraints

When the firms cannot communicate among each other (i.e. they adopt a Bayesian-Nash behaviour with respect to the actions of the other firms), then the Revelation Principle applies[6]. In order to characterise the set of attainable outcomes of the cartel, we can restrict ourselves to direct mechanisms that satisfy a set of Bayesian incentive compatibility constraints (called henceforth BIC) in order to obtain the truthful revelation of the private information held by each firm. But the cartel manager must also ensure that the firms are willing to participate in the cartel: this gives another set of constraints, the Bayesian individual rationality constraints (called BIR).

6.3.1 The Bayesian incentive constraints

Let us call $\mathbf{E}_j\{t(j)\} = \sum_{j=0}^{n} \binom{n}{j} \mu^j (1-\mu)^{n-j} t(j)$ the expected transfer to be given by the efficient firms to the inefficient firms. For a symmetric cartel, we can represent the Bayesian incentive constraints of the two types of firms simply by one lower bound and one upper bound on this expected transfer. Obviously, the lower bound refers to the Bayesian incentive compatibility constraint of an inefficient firm: it gives the minimum expected transfer that the inefficient firms must receive to truthfully reveal their information to the cartel manager. In the case of an efficient cartel, this constraint is[7]

$$BIC(\overline{c}) : \mathbf{E}_j\{t(j)\} \geq n\mu(1-\mu)[\frac{1-(1-\mu)^n}{\mu n}(\pi^m(\underline{c}) - \Delta c q^m(\underline{c})) - \frac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c})].$$

Note that when $\pi^m(\underline{c}) - \Delta c q^m(\underline{c}) \leq 0$, an inefficient firm has no interest to announce it is efficient, because it derives a negative gain from producing the low cost monopoly quantity and it must also pay some transfers to the firms that do announce to be inefficient. Hence, in this case there is no problem of information revelation from the point of view of an inefficient firm. Therefore, we will restrict ourselves to the case in which $\pi^m(\underline{c}) - \Delta c q^m(\underline{c}) \geq 0$.

[6] For the Revelation Principle, see Green and Laffont (1977) and Myerson (1979), among others.
[7] See Appendix A.6.2.

In a similar fashion, the incentive constraint of an efficient firm gives an upper limit to the expected transfer to be given by the efficient firms:

$$BIC(\underline{c}) : \mathbf{E}_j\{t(j)\} \leq n\mu(1-\mu)[\frac{1-(1-\mu)^n}{\mu n}\pi^m(\underline{c}) - \frac{1}{n}(1-\mu)^{n-1}(\pi^m(\overline{c}) + \Delta c q^m(\overline{c}))].$$

6.3.2 The participation constraints

Following Kihlstrom and Vives (1989, 1992), we suppose that if at least one firm refuses to participate in the cartel then the cartel breaks down and the firms play a standard Cournot competition game under asymmetric information. At this stage, we must be explicit about the beliefs of each firm following the rejection of the cartel contract. We assume that if firm i does not accept the cartel contract, this firm will not change its beliefs on the types of the $n-1$ remaining firms, nor will the other firms change their beliefs on the type of the firm that has refused the cartel contract. This is the assumption of passive beliefs as made by Kihlstrom and Vives[8].

In a symmetric and efficient cartel, we can also express the Bayesian individual rationality constraints of an efficient and an inefficient firm as bounds on the expected transfers[9],

$$BIR(\underline{c}) : \mathbf{E}_j\{t(j)\} \leq n\mu[\frac{1-(1-\mu)^n}{\mu n}\pi^m(\underline{c}) - \mathbf{E}\{\pi_{\text{Cournot}}(\underline{c})\}$$

and

$$BIR(\overline{c}) : \mathbf{E}_j\{t(j)\} \geq n(1-\mu)[\mathbf{E}\{\pi_{\text{Cournot}}(\overline{c})\} - \frac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c})].$$

[8] Whether the assumption of passive beliefs is crucial remains an open question in this model with two types. If efficient firms accept the symmetric cartel contract proposed by the cartel manager but the cartel breaks down, it must be the case that an inefficient firm has rejected this contract. The efficient firms know now that there is at least one inefficient firm in the market. This should increase the competitive pressure for the inefficient firms (as the efficient firms will be producing more in comparison with the prior situation) and therefore relax their participation constraints (cf. Cramton and Palfrey, 1995). The reverse holds for an efficient firm: its participation constraint is in fact tighter than the one implied by passive beliefs. We plan to explore this issue in further research. In any case, the assumption is necessary in order to have a tractable analysis of collusion at the level of the subcoalitions.

[9] For the Cournot equilibrium we must distinguish between two cases: one in which both efficient and inefficient firms are producing and one in which only efficient firms produce. Indeed, if the probability of being efficient (μ), the number of firms in the market (n), or the cost differential (Δc) is large, an inefficient firm will anticipate a high competitive pressure and will prefer not to produce at the Cournot equilibrium under incomplete information. In that case, only efficient firms will be active in the market and make profits equal to $(1-\underline{c})^2/(2+(n-1)\mu)^2$. If the inefficient firms also produce at the Cournot equilibrium, the profit of a c-type firm is given by $\frac{1}{4}[1-c-\frac{n-1}{n+1}(1-\mathbf{E}\{c\})]^2$ where $\mathbf{E}\{c\} = \mu\underline{c} + (1-\mu)\overline{c}$ is the expected cost. See Appendix A.6.1 for details.

where $\mathbf{E}\{\pi_{\text{Cournot}}(c)\}$ is the expected profit for a c-type firm under Cournot competition based on incomplete information.

6.3.3 Individual implementability

¿From now on we will use the following terminology.

Definition 6.2 *An efficient cartel is individually implementable if there exists a set of transfers such that all the individual constraints (participation and incentive compatibility) are satisfied*[10].

An illustration of the individual incentive compatibility constraints and the participation constraints, as defined previously, is given in Figure 6.1 for the case of three firms.

The fact that in the present context (with a finite number of efficiency types, the probability distribution of which is i.i.d.), budget-balanced transfers can be found that satisfy the Bayesian incentive compatibility constraints for both types, is in fact just an illustration of the more general result of d'Aspremont and Gérard-Varet (1979) in this respect. The fact that it is possible to find transfers that satisfy the participation constraints for both types (the upperbound of $BIR(\underline{c})$ is above the lowerbound of $BIR(\bar{c})$) is not very surprising either: as the cartel profit exceeds the competitive industry profit in each state of nature, it will be possible to form the cartel, given that every firm tells the truth.

Comparing the lowerbound of $BIC(\bar{c})$ with the upperbound of $BIR(\underline{c})$ and the lowerbound of $BIR(\bar{c})$ with the upperbound of $BIC(\underline{c})$ is less straightforward. In particular the first comparison poses analytical problems. Nonetheless, we obtain the following proposition.

Proposition 6.1 *Sufficient conditions for the cartel to be individually implementable are that (i) $\Delta c \leq \Delta c_{ES}$, where $\Delta c_{ES} = \Delta c_{ES}(\mu, n)$ is the largest cost difference for which efficient firms are still willing to share the cartel profits equally with the inefficient firms and (ii) $\Delta c \geq \Delta c^0_{BIC}$, where $\Delta c^0_{BIC} = \Delta c^0_{BIC}(\mu, n)$ is the level of Δc for*

[10] Our definition concerning the implementability of an efficient cartel includes the incentive constraints and the individual participation constraints. Usually, the term 'implementability' refers only to the problem of information revelation. But we use this terminology to make a clear-cut distinction between the individual constraints and the constraints concerning the subcoalitions that we will define later on.

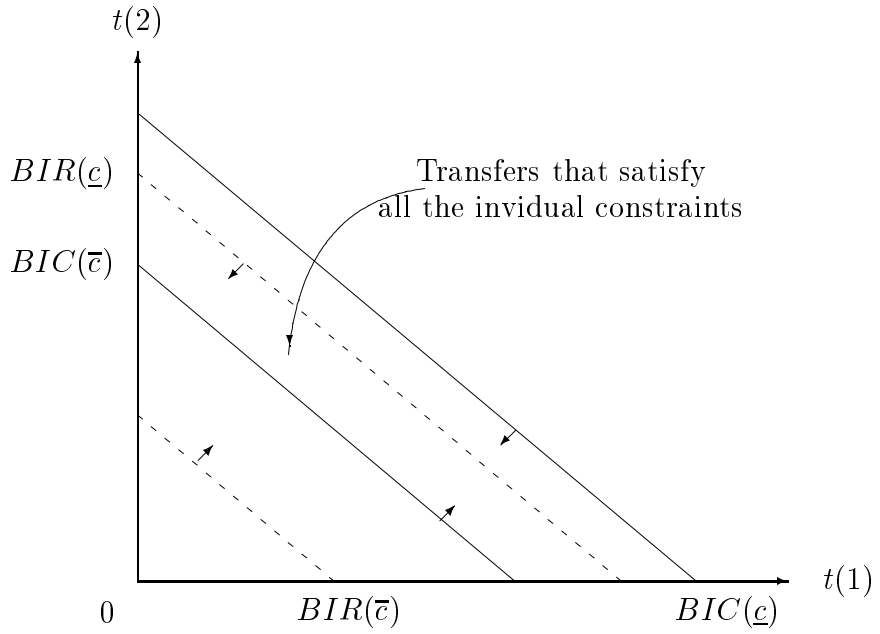

Figure 6.1: Individual implementation of an efficient cartel is always possible.

which, in the absence of transfers, the inefficient firms are indifferent between lying and truthtelling.

Proof. See Appendix A.6.3.

Under the first condition, low cost firms are effectively willing to share the profits from the cartel equally with the high cost firms. Under this transfer scheme, the 'equal sharing rule', high cost firms have no incentive to misrepresent themselves, so that $BIC(\overline{c})$ is satisfied. This sufficient condition obviously amounts to cost differences which are small. The second condition stems from the fact that $BIC(\overline{c})$ can already be satisfied by *zero* transfers, when the cost differences are large. In that case the bound imposed by $BIC(\overline{c})$ (being zero) is clearly below the bound of $BIR(\underline{c})$ (positive).

As in Cramton and Palfrey (1990), it turns out to be impossible to completely compare the two constraints $BIR(\underline{c})$ and $BIC(\overline{c})$ for general parameters without using simulations. For cost differences inbetween 'small' and 'large', we resort to simple simulations that show that the cartel is also implementable for these parameter values[11]. The result is in line with the findings of Kihlstrom and Vives (1989, 1992)

for the polar cases of two firms and a continuum of firms.

Note that there exists in fact a whole region (of non-zero measure) of transfers that enable the cartel manager to simultaneously satisfy the Bayesian incentive constraints and the Bayesian participation constraints for both types of firms. These constraints can be binding or not, depending on the profile of transfers chosen.

6.3.4 Individual implementability with minimal transfers

Having established the result that an efficient cartel can be implemented, it is interesting to focus on a few characteristics of the possible cartels. One such characteristic is the minimal transfers that are necessary to operate it. After all, a cartel that operates in an industry that is surveyed by an antitrust authority will normally, in order to minimize the risk of detection, have a certain preference for cartel schemes that involve as few transfers as possible. We find, on the basis of a sufficient condition and simulations, that whenever positive expected transfers are necessary to implement the efficient cartel, it is the Bayesian incentive compatibility constraint of an inefficient firm that determines the level of the necessary expected transfers[12].

In terms of Figure 6.1, this result tells us that whenever positive transfers are needed, the lowerbound implied by $BIC(\bar{c})$ is above the lowerbound implied by $BIR(\bar{c})$. Note that this is fairly intuitive: when the cost difference is almost zero ($\bar{c} \approx \underline{c}$), high transfers will be necessary to refrain a 'high cost' firm from telling that it is low cost; in fact, a high cost firm will require nothing less than equal sharing of the cartel profits among all firms. On the other hand, $BIR(\bar{c})$ implies that a high cost firm gets more than what it would get under Cournot competition under (almost) complete information and (almost) identical costs. Obviously, the equal sharing payoff exceeds this profit: hence, for $\bar{c} \approx \underline{c}$, the lowerbound of $BIC(\bar{c})$ is above the one of $BIR(\bar{c})$. When the cost difference increases, the cost of lying increases and the lowerbound of $BIC(\bar{c})$ goes down – but so will the lowerbound of $BIR(\bar{c})$, as the outside option of Cournot competition becomes less attractive.

It follows that, when positive, the minimal required expected transfer is equal to

$$n\mu(1-\mu)[\frac{1-(1-\mu)^n}{\mu n}(\pi^m(\underline{c}) - \Delta c q^m(\underline{c})) - \frac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c})]$$

[11] The simulations are done after first having reduced the relevant inequality (which is an inequality in four parameters: $\underline{c}$, $\overline{c}$, μ and n), into an inequality with only two parameters (μ and n); see Appendix A.6.3 for the details.

[12] Cf. Appendix A.6.4.

i.e. the right hand side of the Bayesian incentive compatibility constraint for an inefficient firm given above. The influence of the industry parameters on the minimal required transfer can be inferred from this value: for given n, the larger the cost difference (Δc) and the smaller the probability of an other firm being efficient (μ), the smaller the expected transfers can be. The intuition of this result is straightforward. Given that an inefficient firm only gets to produce when a $\bar{c}$-cartel is formed, the gain of truthtelling increases with the probability that this cartel actually forms. This probability is equal to the probability that there is *no* efficient firm around, $(1-\mu)^n$. Hence, the smaller the probability that the other firms are efficient, the better it is for a firm that announces to be inefficient. As regards the cost difference Δc, the larger this difference, the less attractive the option of lying becomes to the inefficient firm. An inefficient firm, if it announces to be efficient, will be in a position to produce but this will be all the more costly, the higher its unit cost of production is relative to the cost level of the efficient firms in the cartel.

Is it possible for a cartel to operate without the use of any transfers at all? For this to be possible it must be the case that the transfer scheme $t(j) = 0, \forall j$ (the 'zero transfer rule') satisfies the incentive and participation constraints for both efficient and inefficient firms. Obviously, the efficient firms will be very contented with the zero transfer rule: they get to produce but do not have to give any of their profits to the firms announcing to be inefficient. Both the incentive constraint and the participation constraint of the efficient type are satisfied. With respect to the inefficient firms, we know from the previous discussion that we only need to see whether the zero transfer rule satisfies the corresponding incentive compatibility constraint: if incentive compatibility is ascertained for an inefficient firm under the zero transfer rule, then it must be the case that the cost difference is quite large. But then the expected profits at the Cournot-Nash equilibrium are that small that participation by the inefficient firm is ensured as well. It follows that transfers are not necessary to operate the cartel when the cost difference is sufficiently large or when the probability of an other firm being efficient, is sufficiently small.

6.4 The formation of subcoalitions

As we have seen in the previous section, there plenty of leeway for the cartel manager to implement the monopoly outcome. Indeed, there is a whole region of transfers that can be used for this purpose (or a hyperplane, in case the cartel manager prefers

to use minimal transfers). However, depending on the chosen transfers, it may be worthwhile for some subcoalitions of firms to come together and try to manipulate the cartel outcome, in order to reach higher gains from participation in the cartel.

6.4.1 The stake of collusion: a heuristic presentation

Assume, as a preliminary step to the forthcoming analysis, that a subcoalition of two inefficient firms can overcome their informational problem and can form (for example, this is the case if the firms in the subcoalition can credibly disclose their private information to each other or if they have a technology to communicate with each other). Suppose that this subcoalition is also able to use side transfers between its members and that it tries to maximize the gains of its members. Considering again the example of three firms, the expected total gain of the subcoalition composed of, say, two inefficient firms when they reveal truthfully their costs to the cartel manager is $\mu t(1)+(1-\mu)\frac{1}{3}\pi^m(\overline{c})$. If these two inefficient firms manipulate their announcements and claim to the cartel manager (who does not know their cost information) that they are composed of one efficient and one inefficient firm then the expected gain of the subcoalition becomes $\mu[\frac{1}{2}(\pi^m(\underline{c})-\Delta cq^m(\underline{c}))+\frac{1}{2}t(2)]+(1-\mu)[\pi^m(\underline{c})-\Delta cq^m(\underline{c})-\frac{1}{2}t(1)]$. In the case where the cartel manager chooses a set of transfers with a high $t(2)$ and a low $t(1)$, the subcoalition of two inefficient firms will have interest to coordinate their cost messages and to lie to the cartel manager. This leads to some inefficient production of the cartel output, which is not desirable from the point of view of the cartel manager. Requiring that the subcoalition of two ('2') firms reveals truthfully that it consists of zero ('0') efficient firms instead of announcing to contain one ('1') efficient firm gives a collusion-proofness constraint, $CPC^2(0,1)$. For example, in Figure 6.2, points A and B enable the cartel manager to implement individually an efficient cartel but do not resist to the formation of active subcoalitions. Note that there are four other constraints preventing deviations of the remaining subcoalitions. This highlights an interesting property of the model we study. Contrary to the previous models on collusion under asymmetric information, there can be many different stakes of collusion depending on the transfers that are used by the cartel manager to individually implement an efficient cartel.

In the next subsection, we model how collusion can take place. We depart from the crude modelling device used in the previous example, where firms can freely disclose their cost information, and explicitly take into account the informational problem at

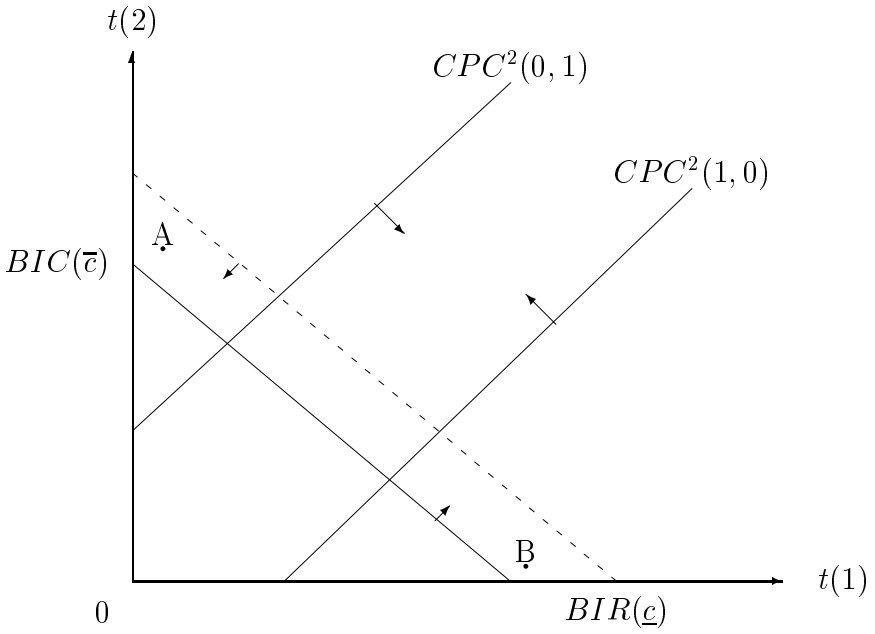

Figure 6.2: The stake of collusion with three firms.

the coalition level.

6.4.2 Subcoalition formation under incomplete information

In this subsection, we propose to use the framework of Laffont and Martimort (1997, 2000) to model the formation of subcoalitions within the cartel. We first begin with the timing that explains how and when a third party can propose to a given subcoalition of k firms a side mechanism aiming at manipulating the reports to be sent by the firms in the cartel contract proposed by the cartel manager.

¿From now on, we will use the following notation. We denote S for the set of indices of the k firms that belong to the given subcoalition and $N\backslash S$ for the set of indices of the $n-k$ other firms that belong to the cartel but not to the subcoalition. Accordingly, c^S (resp. $c^{N\backslash S}$) is the vector of the true cost parameters of the firms belonging to the subcoalition (resp. of the firms belonging to the cartel but not to the subcoalition). The timing is as follows:

1. Nature draws $c^i \in \{\underline{c}, \overline{c}\}$, the private information of firm $i \in N$ according to the common knowledge probability distribution $\{\mu, 1-\mu\}$. Each firm

only learns its own type.

2. The cartel manager proposes all the firms a cartel contract, as defined before.

3. Firm i decides to accept or reject the cartel contract, for $i \in N$. If at least one firm decides to reject it, then the cartel breaks down and the firms come back to Cournot competition under asymmetric information.

4. If all the firms accept the cartel contract then a third party proposes a side mechanism to the firms $i \in S$. This mechanism is composed of a manipulation function $\phi : \{\underline{c}, \overline{c}\}^k \longrightarrow \{\underline{c}, \overline{c}\}^k$ of the messages to be sent to the cartel manager by the firms of the subcoalition, and a vector of monetary side transfers $y : \{\underline{c}, \overline{c}\}^k \longrightarrow \Re$ which must be budget-balanced in each state of nature. If one firm refuses the side mechanism, then the cartel contract is played non-cooperatively by the firms of the subcoalition. If all the firms of the subcoalition accept the side mechanism, they report their private information to the third party non-cooperatively. The third party then assigns the messages to be sent to the cartel manager by the firms in the subcoalition and promises to enforce the corresponding side-transfers.

5. Finally, the reports are sent by the firms into the cartel contract, the quantities and transfers proposed by the cartel manager in the cartel contract take place, as well as the side transfers proposed by the third party in the side mechanism.

The objective of the third party is to maximize the sum of the expected gains of the colluding firms. What can the third party achieve, given that he is uninformed about the private information of the firms in the subcoalition? Because the Revelation Principle applies at this stage of the game (each firm of the subcoalition behaves non-cooperatively within the third party framework), there is no loss of generality in restricting the set of side mechanisms which the third party proposes to the set of direct side mechanisms in which each firm of the subcoalition reveals truthfully its own private information[13]. Hence, the third party must satisfy a set of Bayesian incentive compatibility constraints. But the third party must also ensure that the firms of the subcoalition are willing to participate: a set of Bayesian individual rationality

constraints must be verified by the third party as well, where the reservation value of one firm is its gain when the cartel contract is played non-cooperatively.

For a given state of nature c^S, we denote by $\underline{S}$ and $\overline{S}$ the set of efficient and inefficient firms. We find the following optimality conditions for the manipulation function of the third party (cf. Appendix A.6.5):

$$\phi(c^S)^{opt} \in \arg\max_{\phi(c^S)} \mathbf{E}_{c^{N\setminus S}}\{\sum_{i\in\underline{S}} \pi^i(\underline{c},\phi(c^S),c^{N\setminus S}) + \underline{\epsilon}_i \frac{1-\mu}{\mu}\Delta cq^i(\phi(c^S),c^{N\setminus S}) + \sum_{j\in\overline{S}} \pi^j(\overline{c},\phi(c^S),c^{N\setminus S}) - \overline{\epsilon}_j \frac{\mu}{1-\mu}\Delta cq^j(\phi(c^S),c^{N\setminus S})\}.$$

One can observe that unless the variables $\underline{\epsilon}_i$ (relating to efficient firms) and $\overline{\epsilon}_j$ (relating to inefficient firms) are zero, the third party will not be manipulating efficiently. The variables $\underline{\epsilon}_i$ and $\overline{\epsilon}_j$ reflect the fact that there is an asymmetry of information between the members of the subcoalition. The cost of bridging this informational gap, embodied in $\underline{\epsilon}_i$ and $\overline{\epsilon}_j$, can prevent the subcoalition from realizing all the potential gains of collusion. The values of the ϵs depend on whether the constraints in the program of the third party are binding or not[14].

6.5 Collusion-Proofness

Being faced with the possibility of collusion by a group of firms, what is the optimal response for the cartel manager? In related, but different contexts, Laffont and Martimort (1997, 2000) have shown that the so-called *Collusion-Proofness Principle* applies. This principle states that, in characterising the outcomes that a principal can achieve using general mechanisms, there is no loss of generality to restrict the

[13] We assume that each third party takes in its collusive program only the individual incentive constraints into account. For the case of $n = 3$ firms this is correct as all the subcoalitions are composed of just $k = 2$ individual firms. The third party for the two firms has to ensure that each firm reveals its piece of information. However, for the case of $n > 3$ firms we would in fact also have to consider the possibility that a sub-subcoalition could try to cheat on the third party. Clearly this would add more constraints in the program of the third-party, so that his intermediation would at best be as efficient as the outcome implied by the constraints in our model. Our assumption therefore still helps us to determine an upper bound of the gain of the subcoalition.

[14] When a given cartel contract amounts to ϵs being zero, the third party does not face problems having the information revealed and can implement the efficient side contract. However, it would be wrong to identify a situation in which $\epsilon = 0$ with a situation of complete information at the level of the third party. If this were the case, the cartel manager could incorporate this fact in its cartel contract and take advantage of it.

principal to use a mechanism that is collusion-proof. In this chapter, we will use the following definition:

Definition 6.3 *A cartel contract proposed by the cartel manager is collusion-proof with respect to a subcoalition S if and only if the null side-mechanism $\{\phi = Id_{\{\underline{c},\overline{c}\}^k}, \{y^i = 0\}_{i\in S}\}$ is an optimal response for the third party.*

Differently stated, a grand mechanism is collusion-proof when the third party finds it optimal not to distort the announcements sent by the firms in the subcoalition to the cartel manager and when no monetary side-transfers take place between these firms.

Let us assess whether the Collusion-Proofness Principle applies to the current context of cartel formation as well. The common element in the settings of Laffont and Martimort (1997, 2000) is that the principal is, by assumption, only faced with one possible group of agents that is considering to collude, namely the entire group of agents. In this case, the logic that underlies the Collusion-Proofness Principle is very similar to the one of the Revelation Principle. The outcome of any general mechanism that triggers collusion can be replicated by a direct mechanism which neutralizes the incentive to misrepresent from the coalition's point of view. This allows for a clear characterisation of the implementable allocations.

With the possibility of subcoalitions, it is straightforward to see that the Principle also applies if the cartel manager knows the size and identity of a coalition that possibly colludes[15]. By contrast, when the cartel manager does not know which coalition(s) actually form – the setting of the current chapter – things are different. For, if the cartel manager proposes a general cartel contract and one group colludes, a particular allocation will result, whereas if another group colludes, another allocation will result. Furthermore, a mechanism that is collusion-proof with respect to one subcoalition does not need to be collusion-proof with respect to another subcoalition. This implies that the Collusion-Proofness Principle cannot be generalised in the sense

[15] Cf. Appendix A.6.6. Knowing the size and identity of a coalition that possibly colludes is not the same as knowing that it actively colludes. The first part only refers to the coalition being known to possess a technology to collude (i.e. a third party). As to the second part: even if the firms of the subcoalition know whether they have accepted the side contract or not, any attempt by the cartel manager to elicit this private information can be annihilated by the third party which can punish the deviators from the collusive side-mechanism as much as the cartel manager can reward those who reveal that they are colluding. This also implies that the cartel manager has to satisfy the individual constraints.

that the cartel manager can, without loss of generality, restrict attention to direct mechanisms that are collusion-proof with respect to all conceivably active coalitions: it is not guaranteed that such mechanisms replicate the allocation of the general mechanisms[16]. In particular, it may be optimal for the cartel manager to focus on some groups and to tolerate possible collusion by other groups.

Nonetheless, the focus on efficient cartels allows for the formulation of necessary and sufficient conditions for implementability. After all, an efficient cartel can only be implemented if it can be ensured that no subcoalition finds it in its interest to overstate its efficiency (announcing that it contains more efficient firms than is the case in reality) as this would lead to inefficient production[17].

Let us call $CPC^k(l, l')$ the collusion-proofness constraint that prevents a subcoalition of size k composed of l efficient firms to lie and to annouce to the cartel manager that it is composed of l' efficient firms.

In the next section we will exhibit cartel contracts that implement the efficient cartel. In fact, these cartel contracts turn out to prevent *all* conceivable types of collusion by coalitions (that is, also collusion in the form of announcing to consist of more inefficient firms than is really the case). Whether there exist cartel contracts which satisfy the necessary and sufficient conditions for the implementation of an efficient cartel, but which do not prevent all types of collusion, remains an open question.

6.6 Implementability when collusion is possible

In this section, we ask the following question: is it possible to find a set of transfers such that an efficient cartel is implementable even in the presence of colluding coalitions? For expositional purposes, we will first address the question when it is not necessary at all to use transfers.

[16] This applies not only when the size k is unknown but also when k were known, but not the exact identity of the participating firms. Note furthermore that again this information cannot be elicited, not from the firms actually proposed a collusive side contract nor from the firms not being proposed such a contract.

[17] Note that we are considering symmetric cartel contracts. This is without loss of generality, given that the cartel manager does not know which coalition is formed.

6.6.1 The zero transfer rule

As has already been said, the zero transfer rule is particularly appealing because it enables a cartel to form without having to proceed to any compensatory transfers. Added to this, it appears that whenever a symmetric and efficient cartel is individually implementable with the zero transfer rule, it also satisfies the relevant collusion-proofness constraints ($CPC^k(l,l')$, with $l < l'$).

The intuition is as follows. Let us start with a subcoalition of size k, consisting of l efficient firms and $(k-l)$ inefficient firms. Consider, first, the consequences (in terms of pay-offs for the subcoalition) of one inefficient firm being instructed by the third party to tell the cartel manager that it is efficient. These consequences can be divided into three parts: the change in revenue for the particular inefficient firm itself, the change for the l efficient firms in the subcoalition and the change for the $(k-l)$ inefficient firms. Now, under the zero transfer rule, inefficient firms never get transfers. Hence, for the inefficient firms it does not matter whether one of them is instructed to lie. For the efficient firms, the consequences are undoubtly negative: as soon as the inefficient firm lies, there will be more 'efficient' firms in the cartel, so that the cartel output (i.e. the monopoly output) must be divided among more firms. In that case their pay-offs will be less than when the inefficient firm tells the truth. Hence, in lying the inefficient firm exerts a negative externality on the efficient firms in the subcoalition. So, we must compare the direct gain of lying for the inefficient firm with the negative externality for the efficient firms in the subcoalition. Let us clarify this comparison by means of the following extreme cases. For each case, we will write down the constraints to be satisfied by the zero transfer rule and see which constraint is the most demanding.

The case of $\mu \to 0$ In this case, the subcoalition will act under the assumption that there will be (almost) no efficient firms among the firms that are not in the subcoalition. Then an efficient cartel can be formed without transfers if

$$\begin{aligned} CPC^k(l,l') \qquad & \frac{l'-l}{l}\Delta cq^m(\underline{c}) \geq 0 \\ CPC^k(0,l') \qquad & \frac{k}{n}\pi^m(\overline{c}) - (\pi^m(\underline{c}) - \Delta cq^m(\underline{c})) \geq 0 \\ BIC(\overline{c}) \qquad & \frac{1}{n}\pi^m(\overline{c}) - (\pi^m(\underline{c}) - \Delta cq^m(\underline{c})) \geq 0. \end{aligned}$$

One can observe that whenever $BIC(\overline{c})$ is satisfied, $CPC^k(0,l')$ is also satisfied; the latter is more easily satisfied because a coalition takes into account the above mentioned externality, whereas a firm acting on its own does not. Furthermore, $CPC^k(l,l')$ is always satisfied: as there are no outside efficient firms, the coalition with $l \geq 1$ efficient firms prefers to have the whole monopoly output produced by its efficient firms only.

The case of $\mu \to 1$ Now, the subcoalition will anticipate that all firms are efficient; it is hence with certainty that the low cost cartel will form. An inefficient firm or a coalition with only inefficient firms will then have a strong incentive to misrepresent. The following constraints are to be met

$$\begin{aligned} CPC^k(l,l') \qquad & \frac{l'-l}{(l'+n-k)(l+n-k)}(-(n-k)\pi^m(\underline{c}) + (l+n-k)\Delta cq^m(\underline{c})) \geq 0 \\ CPC^k(0,l') \qquad & -\frac{l'}{l'+n-k}(\pi^m(\underline{c}) - \Delta cq^m(\underline{c})) \geq 0 \\ BIC(\overline{c}) \qquad & -\frac{1}{n}(\pi^m(\underline{c}) - \Delta cq^m(\underline{c})) \geq 0. \end{aligned}$$

The intuition that a coalition with only inefficient firms has a strong incentive to misrepresent, is reflected by the fact that the gain from lying with respect to $CPC^k(0,l')$ is larger than the gain of lying with respect to $CPC^k(l,l')$. Further, one can note that among the $CPC^k(l,l')$, it depends which of the constraints is most demanding. When k is small, the gain of lying is largest for $l' = k$; when k is large, the cost of lying is smallest for $l' = l+1$.

The above two extreme cases, $\mu \to 0$ (all outside firms are inefficient) and $\mu \to 1$ (all outside firms are efficient) give the result that whenever the zero transfer rule satisfies $BIC(\overline{c})$, it also satisfies all collusion-proofness constraints. This leads us to think that also for intermediate cases (in expectation, some outside firms are efficient, some are not) the same applies. Given the complicated character of the expressions for the general case, we are not able to prove this result in general. Nonetheless, we have

Proposition 6.2 *Suppose that only subcoalitions comprising a limited number of firms ($k \leq 2\sqrt{n}$) can form. Under the zero transfer rule, when an efficient cartel is individually implementable, it is also implementable when collusion is possible.*

Proof. See Appendix A.6.8.

The reason is that only when k is not too large, we are able to rank the constraints $CPC^k(0,l')$ and $CPC^k(l,l')$ and to say that the former implies the latter. This is illustrated by the above exposition of the extreme cases. For most cases, it holds that whenever $CPC^k(0,l')$ is satisfied, $CPC^k(l,l')$ is satisfied as well. But when $\mu \to 0$ and Δc takes on the largest relevant value, the cost of lying is greater with respect to $CPC^k(0,l')$ than for $CPC^k(l,l')$ when k is large[18]. Simulations confirm the intuition that the proposition extends to the case of large k as well.

Note that, under the zero transfer rule, also all $CPC^k(l,l')$ with $l' < l$ are satisfied: an efficient firm has no interest to lie because it always gets a part of the monopoly profit and does not have to give back anything to the inefficient firms[19].

6.6.2 Positive transfers

The intuition obtained from the zero transfer rule carries over to the case where individual implementabilty requires positive transfers. Let us again consider a subcoalition of size k, consisting of l efficient firms and $(k-l)$ inefficient firms, contemplating to announce that it consists of $l+1$ efficient firms. The consequences of this action can as before be divided into three parts: the change in revenue for the particular inefficient firm itself, the change for the l efficient firms in the subcoalition and the change for the $(k-l)$ inefficient firms. Now, under the zero transfer rule, inefficient firms always get the same transfer (namely zero), so that the effect of the manipulation on their pay-offs was, in fact, neutral. This leads us to consider, in the case of positive transfers, transfer structures that have the same characteristic as the zero transfer rule, viz. the transfers to be received by an inefficient firm are equal for all (relevant) states of nature. This is what we call the *partial anonymity property with respect to an inefficient firm.* With partially anonymous transfers, it does not matter for the inefficient firms in the subcoalition whether or not one of them is instructed to lie. For the efficient firms, the consequences of such a manipulation are, however,

[18]This is so, because at this cost level, an inefficient firm that is instructed to lie makes no profit by producing the efficient quantity. Then, for a coalition without any efficient firms, lying simply means earning nothing and giving up its part of the inefficient cartel profit ($\frac{k}{n}\pi^m(\overline{c})$), an amount that is increasing in k. For a coalition with at least one efficient firm among its members, lying means reducing the quantity produced by truly efficient firms and giving up the concommittant profits.

[19]Given that all collusion-proofness constraints are satisfied, $\phi = Id$. Then, the incentive compatibility constraints of the third party program coincide with the incentive compatibility constraints of the cartel manager's program. Hence, if both Bayesian incentive constraints are strictly satisfied, the ϵs can be taken zero.

again negative: as soon as one inefficient firm lies, there will be more 'efficient' firms in the cartel, so that the cartel output (i.e. the monopoly output) must be divided among more firms. So, again, there is a negative externality on the efficient firms in the subcoalition. One is left to compare the direct gain for the inefficient firm with the negative externality exerted on the efficient firms. A sufficient condition for the overall benefit of manipulation to be negative is that transfers are taken such that the direct effect of misrepresentation is zero. This is obtained when the transfer that each (reportedly) efficient firm has to pay equals the benefit $\frac{l}{n-l}\{\pi^m(\underline{c}) - \Delta cq^m(\underline{c})\}$ that an inefficient firm gets when it is lying. We summarize this result in the following proposition:

Proposition 6.3 *A sufficient condition for the cartel to be implementable when collusion is possible is that the (partially anonymous) transfers are taken to be* $t(j) = \frac{n-j}{n}\{\pi^m(\underline{c}) - \Delta cq^m(\underline{c})\}$.

Proof. See Appendix A.6.9.

The rule of the above proposition is, in fact, collusion-proof with respect to all conceivable coalitions and dominant strategy incentive compatible. That a dominant strategy incentive compatible transfer rule is collusion-proof is a feature that this model has in common with other models of collusion, e.g. Laffont and Martimort (1997). Dominant strategy incentive compatibility implies that, without any compensation, it is costly for each firm not to tell the truth. This makes effective collusion more difficult to implement by the third party, in our model even impossible[20]. Still, it is easy to construct dominant strategy incentive compatibile rules that are not collusion-proof.

A dominant strategy incentive compatible rule naturally involves the use of more transfers than a Bayesian incentive compatible rule. It remains, therefore, to consider whether the rule of the above proposition satisfies the participation constraint of the efficient type, the relevant upperbound in this respect (by definition it satisfies the Bayesian incentive compatibility constraint of the efficient type). On the basis of two sufficient conditions and some simulations (cf. Appendix A.6.10), we conclude that the (partially anonymous) dominant strategy incentive compatible transfer scheme $t(j) = \frac{n-j}{n}\{\pi^m(\underline{c}) - \Delta cq^m(\underline{c})\}$ implements the efficient cartel.

[20]Note that with this particular scheme of transfers neither $BIC(\underline{c})$ nor $BIC(\overline{c})$ is binding at the optimum. This implies that none of the incentive constraints in the program of the third party are binding at the optimum, so that the particular ϵs can be taken zero.

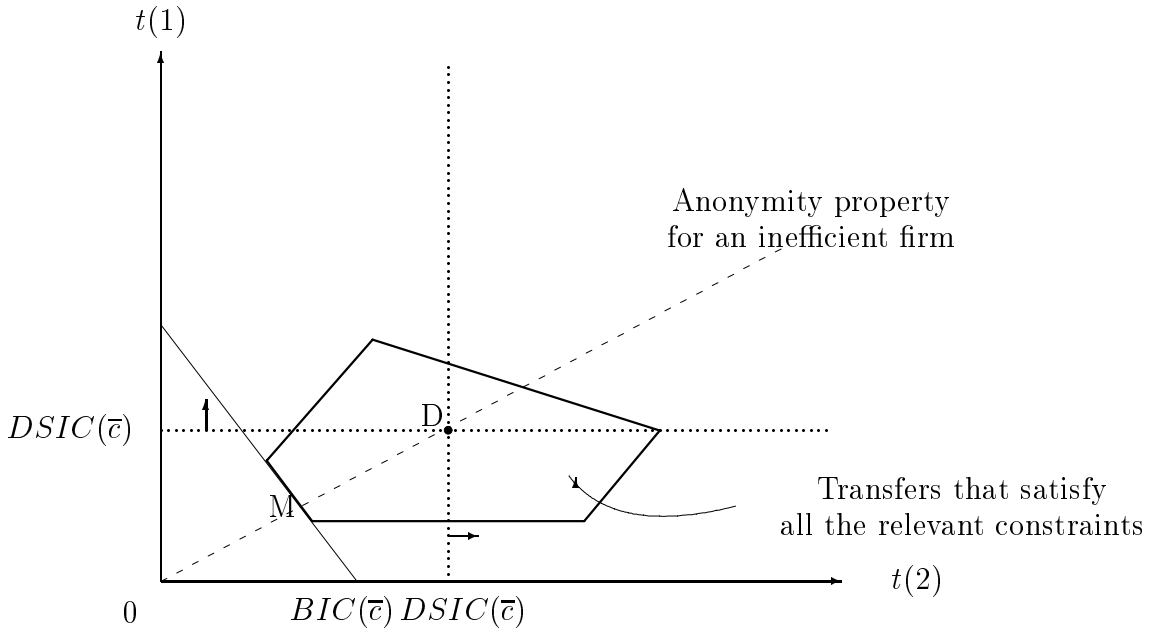

Figure 6.3: The two couples of transfers relating to the existence result.

The rule highlighted above is sufficient to counter collusion, but not necessary; recall that it completely neutralizes the direct gain of the lying firms, so that the overall effect is strictly negative. There certainly exist transfer schemes that leave some gain to the lying firms but still produce an overall negative effect. Simulations indicate that the set of transfers such that the anonymity property for an inefficient firm is satisfied and such that an inefficient firm is individually indifferent between lying and truth-telling to the cartel manager ($BIC(\overline{c})$ is binding) always enables the cartel manager to implement the efficient cartel (whenever positive transfers are necessary). A corollary is that the threat of collusion by subgroups does not seem to change the level of minimally required expected transfers used by the cartel manager to implement an efficient cartel.

In Figure 6.3, two partially anonymous rules (the minimum expected transfer rule, M, and the dominant strategy rule, D) which can implement the efficient cartel in spite of the threat of subcoalition formation are illustrated for the case of a three firm industry.

6.7 Concluding remarks

In this chapter we have addressed the question whether the approach followed in the existing literature to interpret the problem of cartel formation as a classical Principal-Agent problem is not too restrictive. In particular, we have verified whether one basic assumption underlying the use of the Revelation Principle, the assumption about subgroups of firms not attempting to jointly manipulate the cartel in their favour, is determinative for the outcome that a cartel can achieve. The principal result of the chapter is that this is not the case in the context of a two type industry. Whenever a cartel can be implemented with transfers satisfying only the individual incentive and participation requirements of the cartel members, it is also possible to implement this cartel in a collusion-proof way, albeit with a more restricted set of transfers.

Several explanations can be given for this result. First of all, it must be noted that the degree of freedom in the choice of transfers is substantial in this two-type model (cf. d'Aspremont and Gérard-Varet (1979) and Mookherjee and Reichelstein (1992)). This freedom will surely help to find transfers that not only ensure individual implementability, but that also fight collusion. Secondly, there is a strong congruence of interest between the cartel manager and the individual cartel members: after all, the cartel manager is acting costlessly on behalf of the members, by maximizing their total expected profits. As a result, the gain of colluding in a subcoalition is probably relatively limited, compared with situations in which a principal has an objective which goes against the interests of the agents. A third observation is also intuitive: an inefficient firm that is instructed by the third party to represent itself to the cartel manager as being efficient exerts a negative externality on all firms which are truly efficient, including the efficient firms within the subcoalition. If, furthermore, the transfers to the inefficient firms do not depend on the number of inefficient firms present in the cartel and the direct gain of lying for the inefficient firms is not too large, the total effect of the contemplated action is a negative one. In particular, as soon as the individual dominant strategy incentive compatibility constraint of an inefficient firm is satisfied, so will the group incentive compatibility constraints with respect to downward cost announcements.

In our model, collusion appears to have no bite. It is not yet clear to what extent the assumption of independence in the distribution of cost types influences this result. The literature leaves a somewhat mixed impression on this point. In Laffont and Martimort (2000), with two agents and correlated types, collusion by the

grand coalition does matter.

As we have noted, the Collusion-Proofness Principle does not hold when the identities of the possibly colluding subcoalitions are not known to the cartel manager; it no longer needs to be the case that a cartel which is collusion-proof belongs to the class of optimal cartels. Nonetheless, the cartel contract that we have found to be optimal, is collusion-proof also with respect to simultaneously active subcoalitions. This result may be driven by the symmetry of the model[21]. Whether there are cartel contracts that implement the efficient cartel, while letting collusion occur by (some) coalitions, remains an open question[22]. The literature has exhibited many settings in which it is optimal for a principal to let collusion occur. In our context, the basic reason for the Collusion-Proofnesss Principle not to apply can be traced to the fact that, first, the cartel manager does not know the identity of the players he is confronted with (he knows the firms, but not the coalitions) and, secondly, he is not able to elicit this information.

A final remark can be made on the emphasis on the study of efficient cartels. What has been done in this model is to take the production vector of the cartel to be the efficient one (the first best solution) and then to see whether there is a vector of transfers such that all BIR and BIC constraints are satisfied. As far as general transfers are concerned this is not a problem, as it is shown that an efficient cartel is indeed implementable. When transfers cannot be used ('weak cartels' in the meaning of McAfee and McMillan (1992)) it should be interesting to consider cartels which are not necessarily efficient. After all, a standard result of the mechanism design literature is that by distorting 'at the bottom', incentive compatibility constraints can be relaxed. What we could do, therefore, is to write the cartel manager's program in full, giving him also the opportunity to distort the production levels, and to see the impact of collusion in this case. This is a topic for future research.

[21] There is both ex ante symmetry with respect to the cost levels and symmetry with respect to the objective function of the cartel manager.

[22] At this point one may ask the question whether some subcoalitions are more likely to form than others, a question addressed in the literature on endogenous coalition formation (cf. Bloch (1995)). If the cartel manager could 'predict' which coalitions are most likely to form given the cartel scheme it is offering (e.g. because they are the only 'stable' coalitions), it might, in principle, be profitable to focus on these coalitions.

6.8 Appendices

A.6.1. Cournot competition under incomplete information

When no cartel is formed, the firms in the industry engage in Cournot competition under incomplete information. In the Bayesian-Nash equilibrium, a given firm i with cost c^i chooses its production quantity q^i given the expected total production of the other firms, $\mathbf{E}[Q^{N-\{i\}}]$:

$$\max_{q^i}(p(\mathbf{E}\{Q\}) - c^i)q^i = \max_{q^i}(1 - \mathbf{E}[Q^{N-\{i\}}] - q^i - c^i)q^i$$

The first order condition for profit maximization is

$$q^i = \frac{1}{2}(1 - \mathbf{E}[Q^{N-\{i\}}] - c^i) =: R^i(\mathbf{E}[Q^{N-\{i\}}])$$

assuming that $c^i \leq 1 - \mathbf{E}[Q^{N-\{i\}}]$, otherwise firm i will prefer to produce nothing. In a symmetric equilibrium,

$$\mathbf{E}[Q^{N-\{i\}}] = (n-1)\mathbf{E}[q^i] = (n-1)\tfrac{1}{2}(1 - \mathbf{E}[Q^{N-\{i\}}] - \mathbf{E}[c^i])$$

so that

$$\mathbf{E}[Q^{N-\{i\}}] = \tfrac{n-1}{n+1}(1 - \mathbf{E}[c^i])$$

Hence, the equilibrium quantities are

$$q^i_{\text{Cournot}}(c^i) = R^i(\mathbf{E}[Q^{N-\{i\}}]) = \tfrac{1}{2}(1 - c^i - \tfrac{n-1}{n+1}(1 - \mathbf{E}[c^i]))$$

with expected profits

$$\mathbf{E}[\pi^i_{\text{Cournot}}(c^i)] = R^i(\mathbf{E}[Q^{N-\{i\}}]) = \frac{1}{4}(1 - c^i - \frac{n-1}{n+1}(1 - \mathbf{E}[c^i]))^2$$

i.e.

$$\mathbf{E}[\pi^i_{\text{Cournot}}(\underline{c})] = \frac{1}{4(n+1)^2}(2(1-\underline{c}) + (n-1)(1-\mu)\Delta c)^2$$

for a low cost firm and

$$\mathbf{E}[\pi^i_{\text{Cournot}}(\overline{c})] = \frac{1}{4(n+1)^2}(2(1-\overline{c}) - (n-1)\mu\Delta c)^2$$

for a high cost firm. The above expressions hold under the requirement that $c^i \leq 1 - \mathbf{E}\{Q^{N-\{i\}}\}, \forall i$, which translates into $\Delta c \leq \frac{2(1-\underline{c}}{2+(n-1)\mu} =: \Delta c_{\text{limit}}$. For $\Delta c \geq \Delta c_{\text{limit}}$, it is easily verified that only low cost firms will be active in equilibrium (will produce positive quantities) and that their expected profits are equal to $\frac{(1-\underline{c})^2}{(2+(n-1)\mu)^2}$.

A.6.2. Expressing the incentive and participation constraints in terms of bounds on expected total transfers

For a given transfer scheme t, the level of expected total transfers is given by $\mathbf{E}\{t(j)\} = \sum_{j=0}^{n} \binom{n}{j} \mu^j (1-\mu)^{n-j} t(j)$. One can rewrite this level as

$$\begin{aligned}\mathbf{E}\{t(j)\} &= n(1-\mu) \cdot \sum_{j=1}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{n-j} t(j) \\ &= \mathbf{E}\{t(j)\} = n\mu \cdot \sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1} t(j+1).\end{aligned}$$

The Bayesian incentive constraints

If the cartel manager wants to obtain truthful revelation of the private information by the firms he must satisfy

$$BIC(c^i) : \mathbf{E}_{c^{N-\{i\}}}\{\pi^i(c^i, (c^i, c^{N-\{i\}}))\} \geq \mathbf{E}_{c^{N-\{i\}}}\{\pi^i(c^i, (\tilde{c}^i, c^{N-\{i\}}))\} \quad \forall (c^i, \tilde{c}^i) \in \{\underline{c}, \overline{c}\}^2.$$

In the case of a symmetric and efficient cartel, the Bayesian incentive compatibility constraint for an efficient firm is

$$\begin{aligned}BIC(\underline{c}) : \sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1} (\pi^m(\underline{c}) - t(j+1)) \geq \\ \frac{1}{n}(1-\mu)^{n-1}(\pi^m(\overline{c}) + \Delta c q^m(\overline{c})) + \sum_{j=1}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{n-j} t(j).\end{aligned}$$

Observing that $\sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1} t(j+1) = \sum_{j=1}^{n} \binom{n-1}{j-1} \mu^{j-1} (1-\mu)^{n-j} \frac{1}{j} t(j)$ we obtain by putting all the transfer terms $t(j)$ to one side

$$\begin{aligned}\sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1} \pi^m(\underline{c}) + (1-\mu)^{n-1} \frac{1}{n} (\pi^m(\overline{c}) + \Delta c q^m(\overline{c})) \\ \geq \sum_{j=1}^{n-1} \left(\binom{n-1}{j} \frac{1}{n-j} \mu + \binom{n-1}{j-1} \frac{1}{j} (1-\mu) \right) \mu^j (1-\mu)^{n-1-j} t(j).\end{aligned}$$

The right hand side equals $\sum_{j=1}^{n-1} \binom{n-1}{j-1} \mu^{j-1} (1-\mu)^{n-1-j} \frac{1}{j} t(j) = \sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-j} \frac{1}{j+1} t(j+1)$. It follows that an upperbound on the expected total transfer is

given by

$$\mathbf{E}\{t(j)\} = n\mu \cdot \sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1} t(j+1)$$
$$\leq n\mu(1-\mu)[\frac{1}{n\mu}(1-(1-\mu)^n)\pi^m(\underline{c}) - (1-\mu)^{n-1}\frac{1}{n}(\pi^m(\overline{c}) + \Delta cq^m(\overline{c}))]$$

as $\sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1} = \sum_{j=1}^{n} \binom{n-1}{j-1} \mu^{j-1} (1-\mu)^{n-j} \frac{1}{j} = \sum_{j=1}^{n} \binom{n}{j} \mu^{j-1} (1-\mu)^{n-j} \frac{1}{n} = \frac{1}{n\mu}(1-(1-\mu)^n)$.

In a symmetric and efficient cartel, the Bayesian incentive compatibility constraint for an inefficient firm is

$$BIC(\overline{c}) : (1-\mu)^{n-1}\frac{1}{n}\pi^m(\overline{c}) + \sum_{j=1}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{n-j} t(j)) \geq$$
$$\sum_{j=1}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1}(\pi^m(\underline{c}) - \Delta cq^m(\underline{c})) + t(j+1))$$

which, by a similar exercise as above, can be expressed as

$$\mathbf{E}\{t(j)\} \geq n\mu(1-\mu)[\tfrac{1}{n\mu}(1-(1-\mu)^n)(\pi^m(\underline{c}) - \Delta cq^m(\underline{c})) - (1-\mu)^{n-1}\tfrac{1}{n}\pi^m(\overline{c})].$$

The participation constraints

The participation constraints are as follows

$$BIR(c^i) : \mathbf{E}_{c^{N-\{i\}}}[\pi^i(c^i, (c^i, c^{N-\{i\}}))] \geq \mathbf{E}_{c^{N-\{i\}}}\{\pi^i_{\text{Cournot}}(c^i)\}, \forall i \in N.$$

In a symmetric and efficient cartel, the participation constraint for an efficient firm is

$$BIR(\underline{c}) : \sum_{j=0}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{j+1}[\pi^m(\underline{c}) - t(j+1)] \geq \mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}$$

which translates into

$$\mathbf{E}\{t(j)\} \leq n\mu \cdot [\frac{1}{n\mu}(1-(1-\mu)^n)\pi^m(\underline{c}) - \mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}].$$

Similarly, for an inefficient firm

$$BIR(\overline{c}) : \frac{1}{n}(1-\mu)^{n-1}[\pi^m(\overline{c})] + \sum_{j=1}^{n-1} \binom{n-1}{j} \mu^j (1-\mu)^{n-1-j} \frac{1}{n-j} t(j) \geq \mathbf{E}\{\pi_{\text{Cournot}}(\overline{c})\}$$

which becomes

$$\mathbf{E}\{t(j)\} \geq n(1-\mu) \cdot \{\mathbf{E}\{\pi_{\text{ Cournot}}(\overline{c})\} - \tfrac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c})\}.$$

A.6.3. Proof of Proposition 6.1: Individual implementability of an efficient cartel

In order to show that it is always possible to find transfer schemes t that satisfy all four individual constraints, we will compare the upper and lowerbounds on the total expected transfer $\mathbf{E}\{t(j)\}$ which are implied by these constraints.

Comparing $BIC(\underline{c})$ and $BIC(\overline{c})$

The upperbound on $\mathbf{E}\{t(j)\}$ implied by $BIC(\underline{c})$ is above the lowerbound on $\mathbf{E}\{t(j)\}$ implied by $BIC(\underline{c})$ if and only if

$$\frac{1}{n\mu}(1-(1-\mu)^n)\pi^m(\underline{c})+(1-\mu)^{n-1}\frac{1}{n}(\pi^m(\overline{c})+\Delta cq^m(\overline{c}))\geq$$
$$\frac{1}{n\mu}(1-(1-\mu)^n)(\pi^m(\underline{c})-\Delta cq^m(\underline{c}))-(1-\mu)^{n-1}\frac{1}{n}\pi^m(\overline{c})$$

or $(1-(1-\mu)^n)q^m(\underline{c})\geq\mu(1-\mu)^{n-1}q^m(\overline{c})$ which is the case as $1-(1-\mu)^n-\mu(1-\mu)^{n-1}=1-(1-\mu)^{n-1}\geq 0$ and $q^m(\underline{c})\geq q^m(\overline{c})$.

Comparing $BIR(\underline{c})$ and $BIR(\overline{c})$

In order to show that there exist transfers that simultaneously satisfy $BIR(\underline{c})$ and $BIR(\overline{c})$ it will prove to be useful to write the expected Cournot profits $\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}$ and $\mathbf{E}\{\pi^i_{\text{Cournot}}(\overline{c})\}$ using the summation operator: this will facilitate the making of term-by-term comparisons. Write $\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}=\sum_{j=0}^{n-1}\binom{n-1}{j}\mu^j(1-\mu)^{n-1-j}\cdot\underline{\pi}^i_{CN}(j+1)$, where $\underline{\pi}^i_{CN}(j+1)$ is the Cournot-Nash profit for a low cost firm when among the $(n-1)$ other firms in the industry, j firms turn out to be low cost. Similarly, express $\mathbf{E}\{\pi^i_{\text{Cournot}}(\overline{c})\}$ as $\sum_{j=0}^{n-1}\binom{n-1}{j}\mu^j(1-\mu)^{n-1-j}\overline{\pi}^i_{CN}(j)$, with $\overline{\pi}^i_{CN}(j)$ the Cournot profit for a high cost firm when there are j low cost firms among the remaining firms. Then the participation constraints for an efficient and inefficient firm can be written as

$$BIR(\underline{c}):\sum_{j=0}^{n-1}\binom{n-1}{j}\mu^j(1-\mu)^{n-1-j}[\frac{1}{j+1}\pi^m(\underline{c})-\frac{1}{j+1}t(j+1)-\underline{\pi}^i_{CN}(j+1)]\geq 0$$

$$BIR(\overline{c}):\frac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c})+\sum_{j=0}^{n-1}\binom{n-1}{j}\mu^j(1-\mu)^{n-1-j}[\frac{1}{n-j}t(j)-\overline{\pi}^i_{CN}(j)]\geq 0$$

Using $\frac{1}{j+1}\binom{n-1}{j} = \frac{1}{n}\binom{n}{j+1}$and $\frac{1}{n-j}\binom{n-1}{j} = \frac{1}{n}\binom{n}{j}$ and rearranging, these conditions become respectively:

$$\sum_{j=1}^{n-1}\frac{1}{n}\binom{n}{j}\mu^{j-1}(1-\mu)^{n-j}t(j) \leq \sum_{j=1}^{n}\binom{n-1}{j-1}\mu^{j-1}(1-\mu)^{n-j}[\frac{1}{j}\pi^m(\underline{c}) - \underline{\pi}^i_{CN}(j)]$$

$$\sum_{j=1}^{n-1}\frac{1}{n}\binom{n}{j}\mu^{j-1}(1-\mu)^{n-j}t(j) \geq \frac{1-\mu}{\mu}\sum_{j=0}^{n-1}\binom{n-1}{j}\mu^{j}(1-\mu)^{n-1-j}[\overline{\pi}^i_{CN}(j)] - \frac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c})$$

Hence, transfers can be found that satisfy the two constraints if

$$\sum_{j=1}^{n}\binom{n-1}{j-1}\mu^{j-1}(1-\mu)^{n-j}[\frac{1}{j}\pi^m(\underline{c}) - \underline{\pi}^i_{CN}(j)] \geq$$
$$\sum_{j=0}^{n-1}\binom{n-1}{j}\mu^{j-1}(1-\mu)^{n-j}[\overline{\pi}^i_{CN}(j)] - \frac{1}{n\mu}(1-\mu)^{n}\pi^m(\overline{c})$$

which becomes

$$\sum_{j=1}^{n-1}\binom{n-1}{j-1}\mu^{j-1}(1-\mu)^{n-j}\frac{1}{j}[\pi^m(\underline{c}) - j\underline{\pi}^i_{CN}(j) - (n-j)\overline{\pi}^i_{CN}(j)] +$$
$$\mu^{n-1}\frac{1}{n}[\pi^m(\underline{c}) - n\underline{\pi}^i_{CN}(n)] \geq \frac{1}{\mu}(1-\mu)^{n}[\overline{\pi}^i_{CN}(0) - \frac{1}{n}\pi^m(\overline{c})].$$

The above condition is met as $\pi^m(\underline{c}) - j\underline{\pi}^i_{CN}(j) - (n-j)\overline{\pi}^i_{CN}(j) \geq 0, \forall j = 0, ..., n$ and $\overline{\pi}^i_{CN}(0) - \frac{1}{n}\pi^m(\overline{c}) \leq 0$.

Comparing $BIC(\underline{c})$ and $BIR(\overline{c})$

After some adaptations, we can write $BIC(\underline{c})$ and $BIR(\overline{c})$ respectively as

$$\sum_{j=1}^{n-1}\frac{1}{n}\binom{n}{j}\mu^{j-1-j}(1-\mu)^{n-j}t(j) =$$
$$\sum_{j=1}^{n}\binom{n-1}{j-1}\mu^{j-1}(1-\mu)^{n-j}\frac{1}{j}\pi^m(\underline{c}) + (1-\mu)^{n-1}\frac{1}{n}(\pi^m(\overline{c}) + \Delta cq^m(\overline{c}))$$

and

$$\sum_{j=1}^{n-1}\frac{1}{n}\tbinom{n}{j}\mu^{j-1-j}(1-\mu)^{n-j}t(j)\geq\sum_{j=0}^{n-1}\tbinom{n-1}{j}\mu^{j-1}(1-\mu)^{n-1-j}[\overline{\pi}^{i}_{CN}(j)]-\frac{1}{n\mu}(1-\mu)^{n-1}\pi^{m}(\overline{c}).$$

We see that the upperbound implied by $BIC(\underline{c})$ is greater than the lowerbound implied by $BIR(\overline{c})$ if

$$\sum_{j=1}^{n}\tbinom{n-1}{j-1}\mu^{j-1}(1-\mu)^{n-j}\frac{1}{j}\pi^{m}(\underline{c})+(1-\mu)^{n-1}\frac{1}{n}(\pi^{m}(\overline{c})+\Delta cq^{m}(\overline{c}))\geq$$
$$\sum_{j=0}^{n-1}\tbinom{n-1}{j}\mu^{j-1}(1-\mu)^{n-1-j}[\overline{\pi}^{i}_{CN}(j)]-\frac{1}{n\mu}(1-\mu)^{n-1}\pi^{m}(\overline{c}).$$

Observe that with respect to the left hand side it holds that

$$\sum_{j=1}^{n}\tbinom{n-1}{j-1}\mu^{j-1}(1-\mu)^{n-j}\frac{1}{j}\pi^{m}(\underline{c})+(1-\mu)^{n-1}\frac{1}{n}(\pi^{m}(\overline{c})+\Delta cq^{m}(\overline{c}))=$$
$$\sum_{j=1}^{n}\tbinom{n}{j}\mu^{j-1}(1-\mu)^{n-j}\frac{1}{n}\pi^{m}(\underline{c})-(1-\mu)^{n-1}\frac{1}{n}\pi^{m}(\overline{c})+[(1-\mu)^{n-1}\frac{1}{n}\pi^{m}(\overline{c})-(1-\mu)^{n-1}\frac{1}{n}(\pi^{m}(\overline{c})+\Delta cq^{m}(\overline{c}))].$$

This amount is larger than $\sum_{j=1}^{n}\binom{n}{j}\mu^{j-1}(1-\mu)^{n-j}\frac{1}{n}\pi^{m}(\underline{c})-(1-\mu)^{n-1}\frac{1}{n}\pi^{m}(\overline{c})=(\frac{1-(1-\mu)^{n}}{\mu}-(1-\mu)^{n-1})\frac{1}{n}\pi^{m}(\underline{c})=\frac{1-(1-\mu)^{n-1}}{\mu}\frac{1}{n}\pi^{m}(\underline{c})$, where we have used that $\pi^{m}(\overline{c})\geq(\pi^{m}(\overline{c})+\Delta cq^{m}(\overline{c}))$. For the left hand side, it holds that

$$\sum_{j=0}^{n-1}\tbinom{n-1}{j}\mu^{j-1}(1-\mu)^{n-1-j}[\overline{\pi}^{i}_{CN}(j)]-\frac{1}{n\mu}(1-\mu)^{n-1}\pi^{m}(\overline{c})$$
$$=\sum_{j=1}^{n-1}\tbinom{n-1}{j}\mu^{j-1}(1-\mu)^{n-1-j}[\overline{\pi}^{i}_{CN}(j)]+\frac{1}{\mu}(1-\mu)^{n-1}\overline{\pi}^{i}_{CN}(0)-\frac{1}{n\mu}(1-\mu)^{n-1}\pi^{m}(\overline{c}).$$

This amount is smaller than $\sum_{j=1}^{n-1}\binom{n-1}{j}\mu^{j-1}(1-\mu)^{n-1-j}[\overline{\pi}^{i}_{CN}(j)]\leq\sum_{j=1}^{n-1}\binom{n-1}{j}\mu^{j-1}(1-\mu)^{n-1-j}[\overline{\pi}^{i}_{CN}(0)]=\frac{1-(1-\mu)^{n-1}}{\mu}\overline{\pi}^{i}_{CN}(0)$, where we have used that $\overline{\pi}^{i}_{CN}(0)\leq\pi^{m}(\overline{c})$ and that $\overline{\pi}^{i}_{CN}(j)\leq\overline{\pi}^{i}_{CN}(0),\forall j=1,...,n-1$. Now, as $\frac{1}{n}\pi^{m}(\underline{c})\geq\overline{\pi}^{i}_{CN}(0)$, we can conclude that the upperbound implied by $BIC(\underline{c})$ is greater than the lowerbound implied by $BIR(\overline{c})$.

Comparing $BIR(\underline{c})$ and $BIC(\overline{c})$ (partially involving simulations)

It is possible to find transfers for which the expected total level is above the lowerbound imposed by $BIC(\overline{c})$ and below the upperbound imposed by $BIR(\underline{c})$ if

$$n\mu(1-\mu)[\frac{1}{n\mu}(1-(1-\mu)^n)(\pi^m(\underline{c})-\Delta cq^m(\underline{c}))-(1-\mu)^{n-1}\frac{1}{n}\pi^m(\overline{c})] \leq$$
$$n\mu\cdot[\frac{1}{n\mu}(1-(1-\mu)^n)\pi^m(\underline{c})-\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}].$$

This inequality reduces to

$$\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\} \leq \frac{1}{n}\pi^m(\underline{c})+\frac{1}{n\mu}(1-\mu-(1-\mu)^n)\Delta cq^m(\underline{c})$$
$$+(1-\mu)^n\frac{1}{n}(\pi^m(\overline{c})-\pi^m(\underline{c})+\Delta cq^m(\underline{c})). \quad (6.1)$$

Note first that it would be wrong to think that always $\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\} \leq \frac{1}{n}\pi^m(\underline{c})$; when a low cost firm has a good probability to be the only efficient firm and the cost difference is fairly large, it may be that its conditionally expected Bayesian-Cournot profit is higher than the monopoly profit divided by all the firms.

Sufficient condition 1 *A sufficient condition for the lowerbound imposed by $BIC(\overline{c})$ to be below the upperbound imposed by $BIR(\underline{c})$ is that $\Delta c \leq \Delta c_{ES}$, where Δc_{ES} is the largest cost difference for which efficient firms are still willing to share the cartel profits equally with the inefficient firms.*

Observe first that the last two terms in inequality (6.1) are both positive (by a revealed preference argument, $\pi^m(\overline{c}) \geq \pi^m(\underline{c})-\Delta cq^m(\underline{c})$). A sufficient condition for inequality (6.1) to be satisfied is that $\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\} \leq \frac{1}{n}\pi^m(\underline{c})$, i.e. low cost firms prefer sharing equally the cartel profit $\pi^m(\underline{c})$ to getting the expected profit out of Cournot competition under incomplete information. Elaborating on this, we obtain

$$\frac{1}{n}\pi^m(\underline{c}) \leq \mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\} \Leftrightarrow \frac{1}{n}\pi^m(\underline{c}) \leq \frac{1}{4(n+1)^2}(2(1-\underline{c})+(n-1)(1-\mu)\Delta c)^2$$
$$\Leftrightarrow \Delta c \leq \frac{(\sqrt{n}-1)^2(1-\underline{c})}{(n-1)\sqrt{n}(1-\mu)} =: \Delta c_{\text{ES}}.$$

The above calculation is based on the Cournot profit for an efficient firm when high costs firms are active in equilibrium (i.e. $\Delta c \leq \Delta c_{\text{limit}}$). Obviously, when $\frac{1}{n}\pi^m(\underline{c}) \leq$

$\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}$ for $\Delta c \leq \Delta c_{\text{limit}}$, it is also the case for larger Δc. Therefore, when the obtained level Δc_{ES} turns out to be larger than Δc_{limit}, we can conclude that $\frac{1}{n}\pi^m(\underline{c}) \leq \mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}$ for all Δc. As a corollary, an alternative sufficient condition is that $n \geq \left(\frac{2-\mu}{\mu}\right)^2$. Indeed, when $\frac{1}{n}\pi^m(\underline{c}) \leq \mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\}$ for $\Delta c \leq \Delta c_{\text{limit}}$, it is also the case for larger Δc. Then, $\Delta c_{\text{ES}} \leq \Delta c_{\text{limit}} \Leftrightarrow \underline{c} + \frac{(\sqrt{n}-1)^2(1-\underline{c})}{(n-1)\sqrt{n}(1-\mu)} \leq \frac{2+(n-1)\mu\underline{c}}{2+(n-1)\mu} \Leftrightarrow n \geq \left(\frac{2-\mu}{\mu}\right)^2$.

Sufficient condition 2 *A sufficient condition for the lowerbound imposed by $BIC(\overline{c})$ to be below the upperbound imposed by $BIR(\underline{c})$ is that $\Delta c \geq \Delta c^0_{BIC}$, where Δc^0_{BIC} is the level of Δc for which, under zero transfers, a high cost firm is indifferent between lying and truthtelling.*
The zero transfer rule satisfies $BIC(\overline{c})$ if

$$0 \geq \frac{1}{n\mu}(1-(1-\mu)^n)(\pi^m(\underline{c}) - \Delta c q^m(\underline{c})) - (1-\mu)^{n-1}\frac{1}{n}\pi^m(\overline{c}).$$

Using the relations $\pi^m(.) = (q^m(.))^2$ and $q^m(\overline{c}) = q^m(\underline{c}) - \frac{1}{2}\Delta c$ this comes down to $\mu(1-\mu)^{n-1}\frac{1}{4}(\Delta c)^2 + (1-(1-\mu)^{n-1})\Delta c q^m(\underline{c}) - (1-(1-\mu)^{n-1})(q^m(\underline{c}))^2 \geq 0$. This relation with equality is quadratric in Δc and has two roots, one negative and one positive. From this, it follows that the inequality is met for $\Delta c \geq \Delta c^0_{BIC}$, where

$$\Delta c^0_{BIC} = \frac{1}{\mu(1-\mu)^{n-1}}\{\sqrt{(1-(1-\mu)^n)(1-(1-\mu)^{n-1})} - (1-(1-\mu)^{n-1})\}(1-\underline{c}).$$

Simulations Rewriting inequality (6.1), we get

$$\alpha(\Delta c)^2 + \beta q^m(\underline{c})\Delta c + \gamma(q^m(\underline{c}))^2 \leq 0, \tag{6.2}$$

where $\alpha = \frac{n(n-1)^2(1-\mu)^2}{4(n+1)^2} - \frac{1}{4}(1-\mu)^n$, $\beta = \frac{2n(n-1)(1-\mu)}{(n+1)^2} - \frac{1}{\mu}(1-(1-\mu)^{n-1})$ and $\gamma = \frac{4n}{(n+1)^2} - 1$. The determinant of relation (6.2) with equality is $(\beta^2 - 4\alpha\gamma)(q^m(\underline{c}))^2 > 0$ as $\gamma < 0$ for $n \geq 2$ and $\alpha \geq 0$. Furthermore, as $(\beta^2 - 4\alpha\gamma)(q^m(\underline{c}))^2 \geq \beta^2(q^m(\underline{c}))^2$, there is one positive and one negative root. Hence, for the bound of $BIC(\overline{c})$ to be below the bound of $BIR(\underline{c})$ one must have that $\overline{c} \leq \overline{c}'$, where

$$\overline{c}' = \underline{c} + \frac{1}{4\alpha}(-\beta + \sqrt{(\beta^2 - 4\alpha\gamma)})(1-\underline{c})$$

We know that if the bound of $BIC(\overline{c})$ is below the bound of $BIR(\underline{c})$ for $\overline{c} \leq \overline{c}_{\text{limit}}$, it also holds for all $\overline{c} \in [\overline{c}_{\text{limit}}, 1]$. Let us compare $\overline{c}'$ and $\overline{c}_{\text{limit}}$; if $\overline{c}' \geq \overline{c}_{\text{limit}}$, we can conclude that the bound of $BIC(\overline{c})$ is below that of $BIR(\underline{c})$ for all parameter

values. Observe that both $\overline{c}'$ and $\overline{c}_{\text{limit}}$ are linear in $\underline{c}$ and that when $\underline{c} = 1$, we have $\overline{c}' = \overline{c}_{\text{limit}} = 1$. We therefore only have to compare $\overline{c}'$ and $\overline{c}_{\text{limit}}$ for $\underline{c} = 0$. Simple simulations on the resulting inequality in two parameters, n and μ, show that indeed $\overline{c}' \geq \overline{c}_{\text{limit}}$; we conclude that for all parameter values, one can find transfers that satisfy $BIC(\overline{c})$ and $BIR(\underline{c})$.

A.6.4. Minimal transfers

The two lowerbounds on the transfers are given by the constraints $BIC(\overline{c})$ and $BIR(\overline{c})$. We have

$$BIC(\overline{c}) \Leftrightarrow \frac{1}{n}\sum_{j=1}^{n-1}\binom{n}{j}\mu^{j-1}(1-\mu)^{n-j-1}t(j) \geq$$
$$-\frac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c}) + \frac{1-(1-\mu)^n}{\mu n}[\pi^m(\underline{c}) - \Delta c q^m(\underline{c})]$$

$$BIR(\overline{c}) \Leftrightarrow \frac{\mu}{n}\sum_{j=1}^{n-1}\binom{n}{j}\mu^{j-1}(1-\mu)^{n-j-1}t(j) \geq -\frac{1}{n}(1-\mu)^{n-1}\pi^m(\overline{c}) + \mathbf{E}\{\pi_{\text{Cournot}}(\overline{c})\}.$$

We will show that, whenever positive transfers are necessary to implement the efficient cartel, it is constraint $BIC(\overline{c})$ that determines the level of necessary expected transfers. This is equivalent to saying that

$$\frac{1-(1-\mu)^n}{n}[\pi^m(\underline{c}) - \Delta c q^m(\underline{c})] \geq \mathbf{E}\{\pi_{\text{Cournot}}(\overline{c})\} - \frac{1}{n}(1-\mu)^n\pi^m(\overline{c}) \tag{6.3}$$

for all parameter values for which $BIR(\overline{c})$ requires positive transfers. The latter requirement amounts to $(-(1-\mu)^{\frac{n-1}{2}} + \frac{2\sqrt{n}}{n+1} + \frac{(n-1)\sqrt{n}}{n+1}\mu)\Delta c \leq 2(-(1-\mu)^{\frac{n-1}{2}} + \frac{2\sqrt{n}}{n+1})q^m(\underline{c})$, i.e. to the requirement $\Delta c \leq \Delta c^0_{BIR}$, where Δc^0_{BIR} is the cost difference for which the zero transfer rule just satisfies $BIR(\overline{c})$

$$\Delta c^0_{BIR} = \frac{2(-(1-\mu)^{\frac{n-1}{2}} + \frac{2\sqrt{n}}{n+1})}{-(1-\mu)^{\frac{n-1}{2}} + \frac{2\sqrt{n}}{n+1} + \frac{(n-1)\sqrt{n}}{n+1}\mu}$$

where we can bear in mind that when $-(1-\mu)^{\frac{n-1}{2}} + \frac{2\sqrt{n}}{n+1} + \frac{(n-1)\sqrt{n}}{n+1}\mu < 0$, i.e. when μ is small enough, the zero transfer rule always satisfies $BIR(\overline{c})$, regardless of the cost difference.

Rewriting inequality (6.3) gives us $\alpha(\Delta c)^2 + \beta q^m(\underline{c})\Delta c + \gamma(q^m(\underline{c}))^2 \leq 0$, where $\alpha = n(2+(n-1)\mu)^2 - (n+1)^2(1-\mu)^n$, $\beta = 4(n-1)(n-1-2n\mu)$ and $\gamma = -4(n-1)^2$. If α is positive, i.e. $n(2+(n-1)\mu)^2 - (n+1)^2(1-\mu)^n > 0$ (μ not too small), then the discriminant Δ is positive and there is one positive and one negative root (which can be discarded). Call the relevant positive root $\Delta\hat{c}_+$. It holds that $BIC(\overline{c})$ requires higher transfers than $BIR(\overline{c})$ if for all $\Delta c \leq \Delta c^0_{BIR}$, we have that $\Delta c \leq \Delta\hat{c}_+$. By comparing $\Delta\hat{c}_+$ and Δc^0_{BIR} for the relevant parameter values, we can see whether this is the case. Observe that both Δc^0_{BIR} and $\Delta\hat{c}_+$ are linear in $\underline{c}$ and that when $\underline{c} = 1$, we have $\Delta c^0_{BIR} = \Delta\hat{c}_+ = 0$. We therefore only have to compare Δc^0_{BIR} and $\Delta\hat{c}_+$ for $\underline{c} = 0$. Simple simulations on the resulting inequality in two parameters, n and μ, show that indeed $\Delta c^0_{BIR} \leq \Delta\hat{c}_+$. When α is negative (μ not too large), it easily follows that $BIR(\overline{c})$ is satisfied by the zero transfer rule, so that also in this case $BIC(\overline{c})$ is necessarily the most demanding. We conclude that for all relevant parameter values, $BIC(\overline{c})$ requires higher transfers than $BIR(\underline{c})$.

A.6.5. The third party's program

The objective of the third party is to maximize the sum of the expected gains of the colluding firms. Suppose the third party proposes to the group of firms a side contract (ϕ, y), where ϕ is the vector of manipulated cost messages of the considered group to be sent to the principal and y the vector of internal side payments.

Let us consider the gain of firm $i \in S$, with cost parameter c^i that annouces $\tilde{c}^i$ to the third party, when the cost reports of the remaining firms in the subcoalition to the third party are $\tilde{c}^{S-\{i\}} \in \{\underline{c}, \overline{c}\}^{k-1}$ and the cost reports of the firms belonging to the cartel but not to the subcoalition to the cartel manager are $\tilde{c}^{N\setminus S} \in \{\underline{c}, \overline{c}\}^{n-k}$. In this case, we denote by $m = (\phi(\tilde{c}^i, \tilde{c}^{S-\{i\}}), \tilde{c}^{N\setminus S})$ the vector formed by the messages sent, on the recommendation of the third party, by the firms of the subcoalition to the cartel manager (that is $\phi(\tilde{c}^i, \tilde{c}^{S-\{i\}})$) and the messages sent by the other firms (that is $\tilde{c}^{N\setminus S}$). The gain of firm i belonging to the subcoalition is then equal to

$$\pi^i_S(c^i, m) = [P(Q(m)) - c^i]q^i(m) - t^i(m) - y^i(\tilde{c}^i, \tilde{c}^{S-\{i\}})$$

The expected gain of firm i is $\mathbf{E}_{\tilde{c}^{N\setminus S}}\mathbf{E}_{\tilde{c}^{S-\{i\}}}[\pi^i_S(c^i, m)]$, where, for expositional purposes, we have decomposed the total expectation in the expectation over the private information of the firms that do not belong to the subcoalition and the expectation

over the private information of the firms that belong to the subcoalition other than firm i.

Because the Revelation Principle applies at the third party's level, in order to ensure the revelation of information and the participation of the firms in the subcoalition, the third party must solve the following problem

$$\max_{\{\phi,y\}} \mathbf{E}_{c^{N\setminus S}} \mathbf{E}_{c^S} \{ \sum_{i\in S} \pi^i(c^i, \phi(c^i, c^{S-\{i\}}), c^{N\setminus S}) \}$$

subject to

$$BIC^{TP}(c^i) \; : \; \mathbf{E}_{c^{N\setminus S}} \mathbf{E}_{c^{S-\{i\}}} \pi_S^i(c^i, \phi(c^i, c^{S-\{i\}}), c^{N\setminus S}) \geq$$
$$\mathbf{E}_{c^{N\setminus S}} \mathbf{E}_{c^{S-\{i\}}} \pi_S^i(c^i, \phi(\tilde{c}^i, c^{S-\{i\}}), c^{N\setminus S}), \; \forall (c^i, \tilde{c}^i) \in \{\underline{c}, \overline{c}\}^2 \, \forall i \in N$$
$$BIR^{TP}(c^i) \; : \; \mathbf{E}_{c^{N\setminus S}} \mathbf{E}_{c^{S-\{i\}}} \pi_S^i(c^i, \phi(c^i, c^{S-\{i\}}), c^{N\setminus S}) \geq \tilde{\Pi}^i(c^i), \; \forall c^i \in \{\underline{c}, \overline{c}\} \, \forall i \in N$$
$$BB^{TP}(c^S) \; : \; \sum_{i\in S} y^i(c^S) = 0, \; \forall c^S \in \{\underline{c}, \overline{c}\}^k$$

where $\tilde{\Pi}^i(c^i)$ is the gain of firm i when it rejects the side-mechanism proposed by the third party and plays non-cooperatively the symmetric cartel contract proposed by the cartel manager with passive beliefs.

Now, let us denote by $\underline{\delta}_i^{TP}$, $\overline{\delta}_i^{TP}$, $\underline{\nu}_i^{TP}$, $\overline{\nu}_i^{TP}$ and $\rho^{TP}(c^S)$ the Lagrange multipliers associated respectively to $BIC_i^{TP}(\underline{c})$, $BIC_i^{TP}(\overline{c})$, $BIR_i^{TP}(\underline{c})$, $BIR_i^{TP}(\overline{c})$ and $BB^{TP}(c^S)$. Optimizing with respect to the side-transfers $y^i(c^S)$ and $y^j(c^s)$, where firm i is efficient and firm j is inefficient we find the following conditions

$$\rho(c^S) = \mu^l (1-\mu)^{k-l} [\underline{\delta}_i^{TP} + \underline{\nu}_i^{TP} - \overline{\delta}_i^{TP}]$$
$$\rho(c^S) = \mu^l (1-\mu)^{k-l} [\overline{\delta}_j^{TP} + \overline{\nu}_j^{TP} - \underline{\delta}_j^{TP}].$$

which results in the following relationship:

$$\underline{\delta}_i^{TP} + \underline{\nu}_i^{TP} - \overline{\delta}_i^{TP} = \overline{\delta}_j^{TP} + \overline{\nu}_j^{TP} - \underline{\delta}_j^{TP}$$

for every efficient firm i and inefficient firm j in the subcoalition. It can be interpreted as follows: suppose that the third party wants to increase marginally the side transfer given by an efficient firm in the corresponding state of nature. The direct benefit of this operation is to relax the budget balance condition. It also enables the third party to relax the incentive constraint of the inefficient firm (receiving the transfer). But it hardens the incentive and participation constraints of an efficient firm. In the same

way, increasing the side transfer corresponding to an inefficient firm has the same kind of consequences. The previous relationship shows that the cost of increasing the side transfer given by an efficient firm must equate the benefit of the corresponding increase in the side transfer received by an inefficient firm. Hence, there is no further need to modify these side transfers.

For a given state of nature c^S we denote $\underline{S}$ and $\overline{S}$ the set of efficient and inefficient firms. After manipulations the optimality conditions for the manipulation function of the third party are

$$\phi(c^S)^{opt} \in \arg\max_{\phi(c^S)} \mathbf{E}_{\tilde{c}^{N\setminus S}}\{\sum_{i\in\underline{S}} \pi^i(\underline{c},\phi(c^S),c^{N\setminus S}) + \underline{\epsilon}_i \frac{1-\mu}{\mu}\Delta cq^i(\phi(c^S),c^{N\setminus S}) + \sum_{j\in\overline{S}} \pi^j(\overline{c},\phi(c^S),c^{N\setminus S}) - \overline{\epsilon}_j \frac{\mu}{1-\mu}\Delta cq^j(\phi(c^S),c^{N\setminus S})\}$$

where $\underline{\epsilon}_i = \frac{\overline{\delta}_i^{TP}}{1+\underline{\delta}_i^{TP}+\underline{\nu}_i^{TP}-\overline{\delta}_i^{TP}}$ and $\overline{\epsilon}_j = \frac{\underline{\delta}_j^{TP}}{1+\overline{\delta}_j^{TP}+\overline{\nu}_j^{TP}-\underline{\delta}_j^{TP}}$. These two variables reflect the fact that there is an asymmetry of information between the members of the subcoalition. In order to bridge this informational gap, the third party must satisfy some constraints. These constraints have some costs, embodied in $\underline{\epsilon}_i$ and $\overline{\epsilon}_j$, that can prevent the subcoalition from realizing all the gains of the collusion. Contrary to Laffont and Martimort (1997, 2000) we must take into account the incentive constraints of both the efficient and the inefficient firms in the program of the third party.

Note that $\underline{\epsilon}_i$ and $\overline{\epsilon}_j$ can be chosen in $[0,1)$ by the cartel manager. In particular, as long as the Bayesian incentive constraints of the third party program are binding, playing on the transfer payments allows the cartel manager to modify the values of the multipliers associated with the constraints in the program of the third party. Henceforth, he can affect the ϵs[23].

[23]In the framework of Laffont and Martimort (1997, 2000) it is possible to determine the optimal value of $\overline{\epsilon}_j$. In our framework, it is not straightforward to determine $\overline{\epsilon}_j$ and $\underline{\epsilon}_i$. The usual argument goes as follows: let us find out where the stake of collusion is; then, let us identify, in an intuitive manner, the active subcoalitions at the equilibrium; depending on the nature of the strategic interaction between the actions of the firms, guess what will be the active collusion-proofness constraints at the equilibrium; once the relevant collusion-proofness constraints are determined, then, if a particular choice of ϵ turns out to relax all these constraints, adopt this value as being the optimal one from the point of view of the cartel manager. However, as expained above, in our model the stakes of collusion are not a priori defined. They depend precisely on the transfer scheme chosen.

A.6.6. On the Collusion-Proofness Principle

If the cartel manager knows the size and identity of a coalition that possibly colludes, the Collusion-Proofness Principle holds. The logic of the principle follows quite closely that of the Revelation Principle and that of Laffont and Martimort (1997)[24].

Let us assume that the k firms of the subcoalition are the first k firms of the cartel. Let us consider a perfect Bayesian equilibrium of the overall game of a cartel contract offer CC in the presence of subcoalition formation such that a side mechanism $SM \neq SM_0$ is chosen by the third party. CC maps the messages $m = (m^1, \cdots, m^k, \tilde{c}^{k+1}, \cdots, \tilde{c}^n) \in M^1 \times ... \times M^k \times \{\underline{c}, \overline{c}\}^{n-k}$ sent by the firms into an allocation $(q,t) \in \Re_+^n \times \Re^n$. CC maximizes the cartel manager's welfare taking into account the continuation equilibrium of the game of coalition formation. We can restrict the space of messages for the members of the subcoalition to be $\{\underline{c}, \overline{c}\}^k$. SM is a side mechanism which can be taken as being a direct side mechanism mapping $\{\underline{c}, \overline{c}\}^k$ into the set of measures on the messages spaces. SM maximizes the sum of the firms' expected gains subject to Bayesian incentive constraints, budget balance conditions and Bayesian individual rationality constraints $\Pi^i(c^i) \geq \tilde{\Pi}^i(c^i)$ where $\Pi^i(c^i)$ is the expected gain of firm i when CC and SM are played.

Consider now the new cartel contract $CC' = CC \circ SM$. It can be shown that there exists a perfect Bayesian equilibrium of the overall game of cartel contract offer with coalition formation in which the cartel manager offers CC' which is a direct cartel contract from $\{\underline{c}, \overline{c}\}^n$ into the decision space $\Re_+^n \times \Re^n$, the third party offers the null side-mechanism SM_0 and this choice is sustained with passive beliefs.

Because SM solves the third party's program with reservation gains $\tilde{\Pi}^i(c^i)$ the null side-mechanism SM_0 solves the third party's program with reservation gains $\Pi^i(c^i)$. Indeed, suppose that it is not the case. Then there would exist a side mechanism SM' such that the third party can achieve a strictly greater payoff for the subcoalition than with SM. Since by definition $\Pi^i(c^i) \geq \tilde{\Pi}^i(c^i)$ the third party's payoff from offering $SM' \circ SM$ would be strictly greater than that achieved with SM. But this would contradict that SM is optimal when CC is offered. Hence, offering the cartel contract

[24] Laffont and Martimort (1997) do not assume passive beliefs in the event that an agent rejects the side contract. They show a collusion-proofness principle which states that there is no loss of generality in restricting oneself to a collusion-proof contract sustained by out-of-equilibrium passive beliefs: this is one equilibrium of the continuation game. Our approach is different: we assume from the outset that beliefs are passive in order to compute the reservation gain in the Cournot competition game. Then, since beliefs are passive at this stage, it is natural to also assume that they are passive in case of refusal of the side contract offered by the third party.

CC' ensures the cartel manager that there is a perfect Bayesian equilibrium of the overall game sustained with passive beliefs in which SM_0 is optimal from the point of view of the third party.

A.6.7. The collusion-proofness constraints for an efficient cartel

Writing the conditions such that the identity function is the optimal manipulation function for the third party gives us all the collusion-proofness constraints. Note that the identity function is symmetric, so that $\underline{\epsilon}_i = \underline{\epsilon}$ for all efficient firms i and $\overline{\epsilon}_j = \overline{\epsilon}$ for all inefficient firms j.

$$CPC^k(0,l')\,\forall 0 < l' : (1-\mu)^{n-k}\{\frac{k}{n}[\pi^m(\overline{c}) - \overline{\epsilon}\frac{\mu}{1-\mu}\Delta cq^m(\overline{c})] - [\pi^m(\underline{c}) - \Delta cq^m(\underline{c}) - \overline{\epsilon}\frac{\mu}{1-\mu}\Delta cq^m(\underline{c}) - t(l') + \frac{k-l'}{n-l'}t(l')]\} + \sum_{j=1}^{n-k}\alpha_{n-k}(j)\{\frac{k}{n-j}t(j) - \frac{l'}{l'+j}[\pi^m(\underline{c}) - \Delta cq^m(\underline{c}) - \overline{\epsilon}\frac{\mu}{1-\mu}\Delta cq^m(\underline{c}) - t(l'+j)] - \frac{k-l'}{n-l'}t(l'+j)\} \geq 0$$

$$CPC^k(l,0) : \sum_{j=0}^{n-k}\alpha_{n-k}(j)\{\frac{l}{l+j}[\pi^m(\underline{c}) + \underline{\epsilon}\frac{1-\mu}{\mu}\Delta cq^m(\underline{c}) - t(l+j)] + \frac{k-l}{n-l-j}t(l+j)\} - \sum_{j=1}^{n-k}\alpha_{n-k}(j)[\frac{k}{n-j}t(j)] - (1-\mu)^{n-k}\{\frac{k-l}{n}[\pi^m(\overline{c}) - \overline{\epsilon}\frac{\mu}{1-\mu}\Delta cq^m(\overline{c})] + \frac{l}{n}[\pi^m(\overline{c}) + \Delta cq^m(\overline{c}) + \underline{\epsilon}\frac{1-\mu}{\mu}\Delta cq^m(\overline{c})]\} \geq 0$$

$$CPC^k(l,l')\,\forall l > l' : \sum_{j=0}^{n-k}\alpha_{n-k}(j)\{\frac{l}{l+j}[\pi^m(\underline{c}) - t(l+j) + \underline{\epsilon}\frac{1-\mu}{\mu}\Delta cq^m(\underline{c})] + \frac{k-l}{n-l-j}t(l+j) - \frac{l'}{l'+j}[\pi^m(\underline{c}) + \underline{\epsilon}\frac{1-\mu}{\mu}\Delta cq^m(\underline{c}) - t(l'+j)] - \frac{k-l'}{n-l'-j}t(l'+j)\} \geq 0$$

$$CPC^k(l,l')\,\forall l<l':\sum_{j=0}^{n-k}\alpha_{n-k}(j)\{\frac{l}{l+j}[\pi^m(\underline{c})-t(l+j)+\underline{\epsilon}\frac{1-\mu}{\mu}\Delta cq^m(\underline{c})]+$$
$$\frac{k-l}{n-l-j}t(l+j)-\frac{l}{l'+j}[\pi^m(\underline{c})+\underline{\epsilon}\frac{1-\mu}{\mu}\Delta cq^m(\underline{c})-t(l'+j)]-$$
$$\frac{l'-l}{l'+j}[\pi^m(\underline{c})-\Delta cq^m(\underline{c})-\overline{\epsilon}\frac{\mu}{1-\mu}\Delta cq^m(\underline{c})-t(l'+j)]-\frac{k-l'}{n-l'-j}t(l'+j)\}\geq 0$$

where we denote for clarifying purposes $\alpha_{n-k}(j)=\binom{n-k}{j}\mu^j(1-\mu)^{n-k-j}$. A standard revealed preferences argument applied to coalitions (l,l') and (l',l) yields the following relationship

$$\underline{\epsilon}(1-\mu)^2-\overline{\epsilon}\mu^2+\mu(1-\mu)\geq 0$$

This indicates that there is a strong interdependence between $\underline{\epsilon}$ and $\overline{\epsilon}$. If the cartel contract is taken to be collusion-proof, the manipulation function will be the identity function and no distortion of the announcements and no side-transfers occur: $\phi=Id$. Hence, the incentive constraints are the same in the program of the cartel manager and in the program of the third party: if the incentive constraint of an efficient firm is binding and the incentive constraint of an inefficient firm is strictly satisfied in the program of the cartel manager, then it will also be the case in the program of the third party. A necessary condition for $\underline{\epsilon}$ and $\overline{\epsilon}$ to be strictly positive simultaneously is that the two BIC-constraints are simultaneously binding, which is not possible in our setting. Then, either $\underline{\epsilon}=0$ and $\overline{\epsilon}>0$, $\underline{\epsilon}>0$ and $\overline{\epsilon}=0$ or $\underline{\epsilon}=\overline{\epsilon}=0$.

Anticipating the final results: We will show that when positive transfers are needed to implement an efficient cartel, the cartel manager can use transfers that are dominant strategy incentive compatible and satisfy the participation and collusion-proofness constraints. In that case, none of the two Bayesian incentive constraints are binding at the optimum in the program of the cartel manager and therefore $\underline{\epsilon}=\overline{\epsilon}=0$. Also, when the cartel manager implements an efficient cartel with transfers having the minimum expected value: playing on the values of $\underline{\epsilon}$ and $\overline{\epsilon}$ in order to relax or strengthen the collusion-proofness constraints does not help the cartel manager: the minimum expected value is entirely determined by $BIC(\overline{c})$ and having one ϵ strictly positive does not help the cartel manager to reduce the amount of minimum expected transfers used to implement an efficient cartel. For our purpose we can choose the particular values $\underline{\epsilon}=\overline{\epsilon}=0$ in this case.

A.6.8. Proof of Proposition 6.2: A sufficient condition for the zero transfer rule to be collusion-proof

Under the zero transfer rule, coalitions of firms never have interest to report to be less efficient than they are in reality: $CPC^k(l,0) \Leftrightarrow \sum_{j=0}^{n-k} \alpha_{n-k}(j)[\frac{l}{l+j}\pi^m(\underline{c})] \geq \frac{1}{n}(1-\mu)^{n-k}[k\pi^m(\overline{c}) + l\Delta cq^m(\overline{c})]$ where $\alpha_{n-k}(j) = \binom{n-k}{j}\mu^j(1-\mu)^{n-k-j}$. The left hand side of this inequality is a sum of positive terms. For $j=0$ the corresponding term is equal to $(1-\mu)^{n-k}\pi^m(\underline{c})$ which is greater than the right hand side of the inequality. Hence, $CPC^k(l,0)$ is satisfied by the zero transfer rule. For $l > l'$, $CPC^k(l,l') \Leftrightarrow \sum_{j=0}^{n-k} \alpha_{n-k}(j)[\frac{l}{l+j} - \frac{l'}{l'+j}]\pi^m(\underline{c}) \geq 0$, which is obviously satisfied.

Now, we will show that $CPC^k(0,k)$ implies $CPC^k(0,l'), \forall l' < k$. Indeed, $CPC^k(0,l')$ where $l' < k$:

$$\sum_{j=0}^{n-k} \alpha_{n-k}(j)\frac{l'}{l'+j}[\pi^m(\underline{c}) - \Delta cq^m(\underline{c})] \leq (1-\mu)^{n-k}\frac{k}{n}\pi^m(\overline{c}).$$

Obviously, as $\frac{l'}{l'+j}$ is increasing in l', the above constraint is most demanding for $l' = k$. Last, because $(1-\mu)^{n-k}\frac{k}{n}$ is increasing in k, the individual constraint $BIC(\overline{c})$ is the most demanding among the constraints $CPC^k(0,k)$.

It remains to be proved that $CPC^k(0,l')$ implies all the constraints $CPC^k(l,l')$, where $l' > l$. The zero transfer rule satisfies $CPC^k(l,l')$ where $l' > l$ if

$$\sum_{j=0}^{n-k} \alpha_{n-k}(j)[\frac{l}{l+j}\pi^m(\underline{c}) - \frac{l'}{l'+j}\pi^m(\underline{c}) + \frac{l'-l}{l'+j}\Delta cq^m(\underline{c})] \geq 0$$

i.e.

$$\sum_{j=0}^{n-k} \alpha_{n-k}(j)\frac{l'}{l'+j}[\pi^m(\underline{c}) - \Delta cq^m(\underline{c})] \leq \sum_{j=0}^{n-k} \alpha_{n-k}(j)[\frac{l}{l+j}\pi^m(\underline{c}) - \frac{l}{l'+j}\Delta cq^m(\underline{c})].$$

Hence, $CPC^k(0,l')$ is more demanding than $CPC^k(l,l')$ if

$$(1-\mu)^{n-k}\frac{k}{n}\pi^m(\overline{c}) \leq \sum_{j=0}^{n-k} \alpha_{n-k}(j)\{\frac{l}{l+j}\pi^m(\underline{c}) - \frac{l}{l'+j}\Delta cq^m(\underline{c})].$$

As $\pi^m(\underline{c}) - \frac{l}{l'}\Delta cq^m(\underline{c}) \geq (1-\frac{l}{l'})\pi^m(\underline{c}) \geq \frac{1}{l+1}\pi^m(\underline{c})$, and $\pi^m(\overline{c}) = [q^m(\overline{c})]^2 = [q^m(\underline{c}) -$

$\frac{1}{2}\Delta c]^2$ we obtain the following very sufficient condition:

$$(1-\frac{k}{n})q^m(\underline{c})^2-(\frac{l}{l+1}-\frac{k}{n})\Delta cq^m(\underline{c})-\frac{k}{4n}(\Delta c)^2\geq 0.$$

A sufficient condition is then that this inequality is satisfied for $\Delta c = q^m(\underline{c})$, or

$$k(l+1)\leq 4n\,\forall l=1,...,k-1\Leftrightarrow k\leq 2\sqrt{n}.$$

A.6.9. Proof of Proposition 6.3: A sufficient condition for a cartel to be collusion-proof

Consider the transfer scheme $t(j)=\frac{n-j}{n}\{\pi^m(\underline{c})-\Delta cq^m(\underline{c})\}$. It satisfies the collusion-proofness constraints $CPC^k(l,l')$ and $CPC^k(l',l)$, with $l>l'$ if

$$\sum_{j=0}^{n-k}\alpha_{n-k}(j)[(\frac{l'}{l'+j}-\frac{l}{l+j})\pi^m(\underline{c})]\leq$$

$$\sum_{j=0}^{n-k}\alpha_{n-k}(j)[(\frac{k-l}{n-l-j}-\frac{l}{l+j})t(l+j)+(\frac{l'}{l'+j}-\frac{k-l'}{n-l'-j})t(l'+j)]\leq$$

$$\sum_{j=0}^{n-k}\alpha_{n-k}(j)[(\frac{l'}{l'+j}-\frac{l}{l+j})\pi^m(\underline{c})+\frac{1}{l+j}(l-l')\Delta cq^m(\underline{c})].$$

Using the fact that the transfers chosen here are such that $t(l+j)=(n-l-j)t(n-1)$, $nt(n-1)=\pi^m(\underline{c})-\Delta cq^m(\underline{c})$ and that $l>l'$ it is immediately checked that these two inequalities are always satisfied. Indeed, the first inequality becomes equivalent to $\sum_{j=0}^{n-k}\alpha_{n-k}(j)\frac{-j(l-l')}{(l+j)(l'+j)}\pi^m(\underline{c})\leq\sum_{j=0}^{n-k}\alpha_{n-k}(j)[\frac{k(l+j)-ln}{l+j}-\frac{k(l'+j)-l'n}{l'+j}]t(n-1)\Leftrightarrow\sum_{j=0}^{n-k}\alpha_{n-k}(j)\frac{j(l-l')}{(l'+j)(l+j)}\Delta cq^m(\underline{c})\geq 0$, which is obviously satisfied. With the same kind of computations, we find that the second inequality is equivalent to $\sum_{j=0}^{n-k}\alpha_{n-k}(j)\frac{l'(l-l')}{(l+j)(l'+j)}\Delta cq^m(\underline{c})\geq 0$ which is also satisfied under our starting assumptions. Note that the revealed preference argument applied respectively to an efficient and an inefficient firm when a symmetric and efficient cartel is implemented, gives us $\pi^m(\underline{c})\geq\pi^m(\overline{c})+\Delta cq^m(\overline{c})$ and $\pi^m(\overline{c})\geq\pi^m(\underline{c})-\Delta cq^m(\underline{c})$.

Note that the proposed transfer scheme satisfies the requirement of dominant strategy implementation. Under the partial anonymity property, an inefficient firm receives the same transfer in each state of nature: $\frac{1}{n-j}t(j)=\frac{1}{n-j-1}t(j+1)$ or, equivalently,

$t(j) = \frac{n-j}{n-1}t(1)$. Dominant strategy implementation amounts to

$$\frac{1}{j+1}[\pi^m(\underline{c}) - \Delta c q^m(\underline{c})] \leq \frac{1}{n-j}t(j) + \frac{1}{j+1}t(j+1) \leq \frac{1}{j+1}\pi^m(\underline{c}).$$

A.6.10. Implementability of an efficient cartel

We now check whether the collusion-proof transfer scheme $t(j) = \frac{n-j}{n}[\pi^m(\underline{c}) - \Delta c q^m(\underline{c})]$ satisfies the participation constraint of an efficient firm (the relevant upperbound in this respect). The scheme satisfies $BIR(\underline{c})$ if

$$\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\} \leq \frac{1}{n}\pi^m(\underline{c}) + \frac{1-\mu}{n\mu}(1-(1-\mu)^{n-1})\Delta c q^m(\underline{c}). \tag{6.4}$$

Sufficient condition 1 *A sufficient condition for the collusion-proof transfer scheme $t(j) = \frac{n-j}{n}\pi^m(\underline{c}) - \Delta c q^m(\underline{c})$ to satisfy the participation constraint of an efficient firm is that $\Delta c \leq \Delta c_{ES}$, where Δc_{ES} is the largest difference for which efficient firms are still willing to share the cartel profits equally with the inefficient firms.*

A sufficient condition for inequality (6.4) to hold is that $\mathbf{E}\{\pi^i_{\text{Cournot}}(\underline{c})\} \leq \frac{1}{n}\pi^m(\underline{c})$, which gives $\Delta c \leq \Delta c_{\text{ES}}$.

Sufficient condition 2 *A sufficient condition for the collusion-proof transfer scheme $t(j) = \frac{n-j}{n}(\pi^m(\underline{c}) - \Delta c q^m(\underline{c}))$ to satisfy $BIR(\underline{c})$ is that the cost difference is large enough: $\Delta c \geq q^m(\underline{c})$*

Simulations Rewriting inequality (6.4), we get

$$\alpha(\Delta c)^2 + \beta q^m(\underline{c})\Delta c + \gamma(q^m(\underline{c}))^2 \leq 0 \tag{6.5}$$

where $\alpha = n(n-1)^2\mu\,(1-\mu)^2$, $\beta = 8n(n-1)\mu(1-\mu)^n$ and $\gamma = 16n\mu - 4\mu(n+1)^2$. As $\gamma < 0$, the determinant of relation (6.5) with equality is positive. Furthermore, as $(\beta^2 - 4\alpha\gamma)(q^m(\underline{c}))^2 \geq \beta^2(q^m(\underline{c}))^2$, there is one positive and one negative root. Call the positive root Δc^*. Then, for the collusion-proof transfer scheme $t(j) = \frac{n-j}{n}(\pi^m(\underline{c}) - \Delta c q^m(\underline{c})$ to satisfy $BIR(\underline{c})$ it is enough that for all relevant Δc (i.e. $\Delta c \leq \Delta c_{\text{limit}}$), it holds that $\Delta c \leq \Delta c^*$. This will be the case whenever $\Delta c^* \geq \Delta c_{\text{limit}}$. Observe that both Δc^* and Δc_{limit} are linear in $\underline{c}$ and that when $\underline{c} = 1$, we have $\Delta c^* = \Delta c_{\text{limit}} = 0$. We therefore only have to compare Δc^* and Δc_{limit} for $\underline{c} = 0$. Simple simulations on the resulting inequality in two parameters, n and μ, show that

indeed $\Delta c^* \geq \Delta c_{\text{limit}}$.; we conclude that for all parameter values, the collusion-proof transfer scheme $t(j) = \frac{n-j}{n}(\pi^m(\underline{c}) - \Delta c q^m(\underline{c}))$ satisfies $BIR(\underline{c})$.

6.9 References

Bloch, F. (1995) 'Endogenous structures of association in oligopolies', Rand Journal of Economics, 26: 537-56.

Cramton, P. and Palfrey, P. (1990), 'Cartel enforcement with uncertainty about costs', International Economic Review, 31: 17-47.

Cramton, P. and Palfrey, P. (1995), 'Ratifiable mechanisms: learning from disagreement', Games and Economic Behaviour, 10(2): 255-283.

D'Aspremont, C. and Gérard-Varet, L. (1979), 'Incentives and incomplete information', Journal of Public Economics, 11: 25-45.

Eckbo, P. (1976), 'The future of world oil', Ballinger, Cambridge, Massachussets USA.

Green, J. and Laffont, J.-J. (1977), 'Characterization of satisfactory mechanisms for the revelation of preferences for public goods', Econometrica, 45: 427-435.

Kihlstrom, R. and Vives, X. (1989), 'Collusion by asymmetrically informed duopolists', European Journal of Political Economy, 5: 371-402.

Kihlstrom, R. and Vives, X. (1992), 'Collusion by asymmetrically informed firms', Journal of Economics and Management Strategy, 1: 371-396.

Laffont, J.-J. and Martimort,D. (1997), 'Collusion under asymmetric information', Econometrica, 65(4): 875-911.

Laffont, J.-J. and Martimort, D. (2000), 'Mechanism design with collusion and correlation', Econometrica, 68(2): 309.

Myerson, R. (1979), 'Incentive compatibility and the bargaining problem', Econometrica, 47: 61-73.

McAfee, R. and McMillan, J. (1986), 'Bidding rings', American Economic Review, 82: 579-599.

Mookherjee, D. and Reichelstein, S. (1992), 'Dominant strategy implementation of Bayesian incentive compatible allocation rules', Journal of Economic Theory, 56: 378-399.

Phlips, L. (1995), 'Competition policy: a game-theoretic perspective', Cambridge University Press, Cambridge, United Kingdom.

Roberts, K. (1985), 'Cartel behaviour and adverse selection', Journal of Industrial Economics, XXXIII(4): 401-413.

Selten, R. (1973), 'A simple model of imperfect competition where four are few and six are many', International Journal of Game Theory, 2: 141-201.

Summary in Dutch - Samenvatting

Het mededingingsrecht is een rechtsgebied waarin economische begrippen een centrale rol innemen. Veel van de sleutelbegrippen binnen het mededingingsrecht - bijvoorbeeld 'concurrentie', 'beperking van de mededinging', 'anti-concurrentieel effect' - zijn immers begrippen die duidelijk een economische lading hebben.

Het is geen eenvoudige zaak om in de praktijk uit te maken of een bepaalde handelsovereenkomst tussen ondernemingen of een bepaalde commerciële handelswijze goed of slecht is voor de concurrentie. Wanneer bijvoorbeeld twee bedrijven een samenwerkingscontract afsluiten voor de ontwikkeling van een nieuwe technologie, wat zal hiervan het effect zijn op de concurrentie in de betreffende markt? Als de twee bedrijven van bescheiden omvang zijn, dan zal het contract geen kwaad kunnen, juist integendeel. Maar wanneer de twee bedrijven precies elkaars twee grootste concurrenten zijn, dan kan hun samenwerking leiden tot een algehele vermindering van de concurrentie in de markt. Maar waar de grens te trekken tussen de twee gevallen?

Het is alleen door naar de precieze omstandigheden van de zaak te kijken dat steekhoudende conclusies kunnen worden getrokken over de effecten op de concurrentie. Naast een analyse van de mogelijke *effecten* zal dit in veel gevallen ook een analyse vereisen van de onderliggende *beweegredenen* voor bedrijven om dergelijk marktgedrag te vertonen. Immers, daar waar het evalueren en voorspellen van markteffecten geen eensluidend oordeel kan geven, zal een analyse van de onderliggende beweegredenen mogelijk meer licht op de zaak werpen. Wanneer men een bepaalde handelspraktijk aantreft zou men zich bijvoorbeeld de vraag kunnen stellen: waarom handelt dit bedrijf op deze manier? Is een anti-competitieve strategie voor dit bedrijf een rationele strategie of is het waarschijnlijker dat dit bedrijf uit geoorloofde motieven handelt?

Binnen de economische wetenschap houdt het vakgebied Industriële Organisatie zich bezig met de relatie tussen het marktgedrag van ondernemingen, de werking van markten en de marktstructuur. Dit wetenschapsgebied was oorspronkelijk vooral empirisch van aard, met een sterke nadruk op het vinden van empirische wetmatighe-

den tussen allerlei bedrijfstak-kenmerken. Zeer bekend is het zogeheten *Structure-Conduct-Performance* denkmodel. Volgens dit denkmodel bepaalt de marktstructuur (het aantal bedrijven in de markt, de mate van verticale integratie, enz.) het marktgedrag (prijsstelling, investeringen in R&D, marketing-inspanningen, enz.) hetgeen vervolgens leidt tot een zeker marktresultaat (marktefficiëntie, winsten). Niettemin, met zijn nadruk op het vinden van empirische wetmatigheden, wierp dit denkmodel weinig licht op de onderliggende marktwerking en ook niet op de vraag of het geobserveerde marktgedrag wel als rationeel was te beschouwen.

Later heeft de Industriële Organisatie de traditionele empirische onderzoekslijn aangevuld (en in sommige gevallen flink veranderd) met een meer theoretische analyse, die zich ondermeer toespitst op de rationaliteit van geobserveerd marktgedrag. Een belangrijk onderzoeksmiddel voor deze analyse is de speltheorie. Speltheorie bestudeert strategische interactie met behulp van wiskundige modellen. Een speltheoretisch model specificeert de deelnemers aan het spel (bijvoorbeeld bedrijven in een markt of individuen in een organisatie), de informatie waarover zij beschikken, de acties die zij kunnen ondernemen, de timing van deze acties, de uitbetalingen die resulteren bij een bepaalde uitkomst en de preferenties van de deelnemers over de mogelijke uitkomsten. In een dergelijk model worden de deelnemers verondersteld elk hun winst (of meer in het algemeen, hun nutsniveau) te maximaliseren in het licht van de informatie waarover zij beschikken en de verwachte acties van de andere spelers. Het algemeen geaccepteerde oplossingsconcept van dergelijke modellen is het zogenaamde Nash-evenwicht. Dit oplossingsconcept is een evenwicht in die zin dat elke speler zijn winsten maximaliseert en hierbij de tegenacties van de andere spelers correct anticipeert.

Voor het grootste deel bestaat dit proefschrift uit het toepassen van dergelijke speltheoretische methoden bij de bestudering van verticale overeenkomsten, d.w.z. overeenkomsten gesloten door bedrijven die op verschillende niveaus van de bedrijfskolom opereren, zoals producenten en hun wederverkopers. Veel bindingen tussen producenten en distibuteurs gaan veel verder dan simpele markttransacties, het leveren van goederen of diensten tegen een bepaalde prijs. In veel gevallen is er sprake van langetermijn contracten die bepaalde verplichtingen leggen op de handelspartners en daarmee hun handelsvrijheid beperken. Bijvoorbeeld, een producent die een exclusief verkoopsgebied toekent aan een distributeur verbindt zich er doorgaans toe niet aan andere distributeurs in dat gebied te leveren. Dergelijke contractuele verplichtingen worden gewoonlijk 'verticale restricties' of 'verticale beperkingen' genoemd.

Het proefschrift bestaat uit twee gedeelten. Het eerste gedeelte, Hoofdstukken 2 en 3, heeft een overzichtskarakter. Hoofdstuk 2 presenteert en evalueert de economische literatuur over de concurrentie-effecten van verticale restricties. De bedoeling van dit hoofdstuk is een begrip te vormen van de belangrijkste inzichten die de economische theorie heeft opgeleverd op dit gebied. Hoofdstuk 3 analyseert de rol van economische analyse vanuit een juridisch perspectief, door in te gaan op de rol van economische analyse bij de toepassing van de Europese mededingingsregels ten aanzien van verticale beperkingen. Zoals zal blijken uit deze twee hoofdstukken zijn er bepaalde eigenaardigheden aan het Europese mededingingsbeleid, maar kan gezegd worden dat het beleid meer en meer in de richting van een echte afweging van mededingingsbevorderende en mededingingsbeperkende effecten evolueert.

Het tweede gedeelte van dit proefschrift geeft drie concrete toepassingen van speltheoretische analyses. In twee hoofdstukken, Hoofdstuk 3 en Hoofstuk 4, wordt de rationaliteit van een bepaald type verticale beperking, verticale prijsbinding, getest in een aantal specifieke marktomstandigheden. Bij verticale prijsbinding is het zo dat bijv. de producent zijn distributeurs ertoe verplicht niet tegen een lagere prijs te verkopen dan de door hem gespecificeerde. Het volgt uit deze twee hoofdstukken dat de noodzaak voor distributeurs om hun vaste kosten terug te verdienen een bepalende rol speelt in de keuze van de producent om al dan niet verticale prijsbinding te hanteren. Hoofdstuk 6, tenslotte, gaat over kartelvorming in bedrijfstakken waar bedrijven onzeker zijn over elkaars kostenniveaus (en daarmee over elkaars prijsstellingsmotieven). In een specifiek model wordt de wijdverbreide opvatting getest dat kartelvorming waarschijnlijker is in geconcentreerde bedrijfstakken dan in bedrijfstakken waarin veel bedrijven actief zijn.